Journal
of Speech and Hearing
Research

September 1959 VOLUME 2 • NUMBER 3

Effects of Hyperventilation and Tetany on the Speech Fluency of Stutterers and Nonstutterers 203
WENDELL JOHNSON, MARTIN A. YOUNG, ADOLPH L. SAHS, AND GEORGE N. BEDELL

The Latin Square Design in Speech and Hearing Research 216
LEONARD S. FELDT

The Conditioned GSR Auditory Speech Threshold 229
JOSEPH B. CHAIKLIN

The Relation Among Three Selected Auditory Speech Thresholds 237
JOSEPH B. CHAIKLIN

A Phonetic Study of Misarticulation of /r/ 244
JAMES F. CURTIS AND JAMES C. HARDY

Training Parents of Children with Functional Misarticulation 258
RONALD K. SOMMERS, SAMUEL P. SHILLING, CLARA D. PAUL, FLORENCE G. COPETAS, DOLORES C. BOWSER, AND COLETTE J. MC CLINTOCK

Sidetone Delay and Reading Rate, Articulation, and Pitch 266
G. DON DAVIDSON

Lipreading Performance and the Rate of the Speaker 271
VINCENT W. BYERS AND LEWIS LIEBERMAN

Speech Pickup by Contact Microphone at Head and Neck Positions... 277
JOHN C. SNIDECOR, IRVING REHMAN, AND DAVID D. WASHBURN

Voice Quality and Anxiety 282
CHARLES F. DIEHL, RICHARD WHITE, AND KENNETH W. BURK

Experimental Blockage of Phonation by Distorted Sidetone 286
COURTNEY STROMSTA

The American Speech and Hearing Association

OFFICERS

President
George A. Kopp, Ph.D.
Wayne State University

Executive Vice-President
Jack Matthews, Ph.D.
University of Pittsburgh

Vice-President
Miriam D. Pauls, Ph.D.
Johns Hopkins Medical Institute

Editor of the Association
Wendell Johnson, Ph.D.
University of Iowa

OFFICERS-ELECT

President-Elect
Stanley Ainsworth, Ph.D.
University of Georgia

Vice-President-Elect
Jack L. Bangs, Ph.D.
Houston Speech and Hearing Center

COUNCIL

The Officers and the following Councilors:

Jon Eisenson, Ph.D. (1959)
Ira J. Hirsh, Ph.D. (1958-60)
Ruth B. Irwin, Ph.D. (1956-59)
James F. Jerger, Ph.D. (1959-61)
Herold S. Lillywhite, Ph.D. (1958-59)
Hayes A. Newby, Ph.D. (1957-59)
Wilbert L. Pronovost, Ph.D. (1958-60)
Sylvia O. Richardson, M.D. (1957-59)
Dean E. Williams, Ph.D. (1959-61)

EXECUTIVE SECRETARY

Kenneth O. Johnson, Ph.D.

The Journal of Speech and Hearing Research

EDITOR

Dorothy Sherman, Ph.D.

ASSISTANT TO THE EDITOR

Dorothy W. Moeller

ASSISTANT EDITOR

Frederic L. Darley, Ph.D.

STATISTICAL CONSULTANT

Leonard S. Feldt, Ph.D.

ASSOCIATE EDITORS

Oliver Bloodstein, Ph.D.
Arthur S. House, Ph.D.
James F. Jerger, Ph.D.
D. E. Morley, Ph.D.
Hildred Schuell, Ph.D.
William R. Tiffany, Ph.D.
John C. Webster, Ph.D.
Joseph M. Wepman, Ph.D.

DEPARTMENT EDITORS

Ernest H. Henrikson, Ph.D.
Book Reviews

Martin F. Palmer, Sc.D.
Records

BUSINESS MANAGER

Kenneth O. Johnson, Ph.D.

APPLICATIONS FOR MEMBERSHIP SHOULD BE ADDRESSED TO THE EXECUTIVE SECRETARY

Effects of Hyperventilation and Tetany on the Speech Fluency of Stutterers and Nonstutterers

WENDELL JOHNSON

MARTIN A. YOUNG

ADOLPH L. SAHS

GEORGE N. BEDELL

The term 'stuttering' is commonly used to denote a disturbance of speech fluency, and most theories of stuttering involve the assumption that factors responsible for this disturbance are to be sought within either the body or 'the person' of the speaker. Presumably relevant organic factors are assumed to be genetic, or acquired through physiological stress, biochemical imbalance, disease, or injury (*6, 20, 34*). Personality factors, assumed to be conducive to stuttered speech, are accounted for generally by reference to faulty personality structure, emotional trauma, and other conditions variously defined, or undefined (*2, 6, 11, 14, 28, 32, 33*).

All the authors are on the faculty of the University of Iowa. Wendell Johnson (Ph.D., University of Iowa, 1931) is Professor of Speech Pathology and Psychology; Martin A. Young (M.A., University of Iowa, 1958) is Research Associate, Speech Pathology and Audiology; Adolph L. Sahs (M.D., University of Iowa, 1931) is Professor and Head, Department of Neurology; and George N. Bedell (M.D., University of Cincinnati, 1946) is Associate Professor of Internal Medicine. This study was made under a grant from the Easter Seal Foundation.

Alternative theories are those which do not presuppose organic or emotional distinctiveness of the speaker classified as a stutterer, stressing instead the learning of behavior disruptive of otherwise normal speech (*2, 14, 15, 16, 17, 28, 33, 35*). The senior author (*17*) has proposed a theory which stresses the interaction among the perceptual, evaluative, symbolic, and motor reactions of the speaker and also the interaction between the speaker's communicative behavior and the perceptual, evaluative, and overt reactions of his listeners; and from a considerable body of relevant data he has drawn the conclusion that the problem called stuttering originates in an interaction between speaker and listener, in which the decisive factor is a crucial judgment made by the listener of the speaker's nonfluency, rather than the degree of the speaker's nonfluency, per se.

The general theory that stuttering, viewed as a disturbance of speech fluency, is due to organic fault has been advanced in several specific forms. One

of these is the hypothesis that stuttering is a form of tetany, or more specifically, latent tetany (*20, 27, 34*), presumably due to parathyroid deficiency with associated neuromuscular hyperexcitability. In latent tetany, the serum calcium[1] remains just above a critical level, and attacks of tetany may be precipitated by excessive emotion, undue strain, or a 'failure in health' (*1*). Prolonged hyperventilation, or overbreathing, may also result in tetany. Although hyperventilation tetany appears not to be fully understood, it has been asserted that it is due to a combination of factors, including reduced hydrogen ion concentration, reduced calcium ion concentration, and a reduced carbon dioxide tension (*3*).

Early reports of hyperventilation tetany were made by Grant and Goldman (*12*) and Collip and Backus (*5*) in 1920. Brown (*3*) has presented an extensive summary of the research dealing with the physiological effects of hyperventilation, but of more interest and relevance to the present research are those effects of hyperventilation on psychomotor performance. Among the representative findings have been those of Gellhorn (*7, 8, 9, 10*), who reported an increase in nystagmic movements during hyperpnea, reduction in hearing, lengthening of latent period of negative afterimages, and a shift from the 'usualness' in word associations. Rushmer and Bond (*24*) and Scow and others (*26*) both found that performance deteriorated on several types of psychomotor test. Rahn and others (*23*) and Lillehei and Balke (*22*) each found, independently, that performance was little affected until carbon dioxide tension fell below 25 mm Hg. Clinical reports (*21, 22*) have indicated that hyperventilation often results in decreased responsiveness to external stimuli.

In the present study continuous rapid infrared carbon dioxide analysis was the procedure chosen to analyze carbon dioxide tension in the lungs. Limitations of this procedure have been reported on by Collier, Affeldt and Farr (*4*). Their conclusion was that although continuous rapid infrared carbon dioxide analysis is the most convenient (though not more accurate than blood analysis), it and any other method suffer from the fact that carbon dioxide concentration varies constantly with time and space in the lungs. Hence, any one value for carbon dioxide tension is an approximation or average of a constantly changing value.

Problem

If stuttering is a form of latent tetany, and stutterers hover on the brink of some critical imbalance or level, then stutterers could be expected to show a deterioration of speech performance under hyperventilation. Indeed, it has been reported (*22, 23*) that even a moderate drop in alveolar carbon dioxide tension due to hyperventilation results in impaired motor performance for nonstutterers. If stuttering is due to latent or active tetany, it is to be predicted that stutterers would perform more stuttering after moderate hyperventilation. Specifically, the problems posed are as follows: (a) What changes

[1]For 15 stutterers tested by Johnson, Sterns, and Warweg (*19*), the serum calcium was found to be within normal limits. Hill (*13*) came to a general conclusion essentially consonant with this finding after reviewing several other relevant studies.

are to be observed in the speech fluency of stutterers and nonstutterers as a result of hyperventilation tetany, induced by voluntary overbreathing, and of the related physiological side-effects of hyperventilation? (b) Do stutterers and nonstutterers differ in respect to alveolar air carbon dioxide tension, as measured by continuous rapid infrared analysis, before and after speaking, before and after hyperventilation, and after speaking following hyperventilation?

Procedure

Experimental Design. The experimental procedure was designed to yield data relevant to the foregoing questions. The experimental subjects (stutterers) and control subjects (nonstutterers) were tested under similar laboratory conditions by means of the same experimental procedure, with certain exceptions to be noted shortly. This procedure was as follows:

Each subject was escorted to the Pulmonary Function Laboratory[2] and instructed to lie in a supine position on a hospital litter. After a brief statement concerning the nature and purpose of the study was presented, data pertaining to the subject's age and educational status were obtained. The carbon dioxide analyzer[3] was calibrated and the subject shown how to use the mouthpiece. The subject took several preliminary breaths to ascertain whether or not the mouthpiece was in place and the equipment was working properly. A nose clamp was applied to assure mouth breathing. The Sanborn Recorder[4] was turned on and alveolar carbon dioxide tension (alveolar CO_2 pressure) was measured on five or six consecutive breaths. The subject was then asked to sit up and a microphone was placed before him. He was handed a folder containing a 150-word reading passage and told to read this passage aloud twice. A tape recorder, in full view of the subject throughout the entire procedure, was then turned on. After the second reading was completed, the subject was told to lie down and replace the mouthpiece in his mouth. At this time the tape recorder was turned off, the Sanborn Recorder turned on, and a sampling of three or four complete breaths obtained.

Still in a supine position with mouthpiece in place, the subject was given instructions for the hyperventilation procedure to follow, and then hyperventilated at a rate approximating 20 deep breaths per minute. The Sanborn Recorder was turned on at the beginning and end of the hyperventilation procedure, long enough to record at least three complete breaths each time. Close watch was kept on each subject to insure that voluntary overbreathing was carried out as prescribed, with respect to both depth and rate, and also to note the presence of any signs of tetany. Periodically, an attempt was

[2]Department of Internal Medicine, University Hospitals, Iowa City, Iowa.

[3]Liston-Becker Gas Analyzer, Model 11, manufactured by Liston-Becker Instrument Company, Inc., Stamford, Conn.

[4]Sanborn Twin-Viso Recorder, Model 60-1300, manufactured by Sanborn Company, Cambridge, Mass.

made to elicit Chvostek's sign.[5] When this sign was quite positive, in the opinion of the experimenter, or when carpopedal spasms were present, hyperventilation was discontinued.

As soon as possible, usually a matter of seconds, after the hyperventilation procedure had been concluded, the mouthpiece was removed and the subject was asked to read the test passage aloud for the third time. At the completion of this reading, the subject assumed the supine position and replaced the mouthpiece, and a final alveolar CO_2 pressure determination was made. The subject was allowed to rest on the litter until the effects of hyperventilation had disappeared and then was dismissed.

The stuttering group also had an experimental control condition, with a scheduled interval of seven days between the two sessions.[6] In the second session, however, a five-minute rest period was substituted for the hyperventilation procedure, and a different, but comparable, 150-word reading passage was employed. During the rest period, the conditions of the hyperventilation procedure were simulated by having the subject in a supine position on the litter with the mouthpiece in place. The two experimental sessions for the stutterers, hyperventilation and no hyperventilation, and the two reading passages, were counterbalanced, and subjects were assigned at random to one of the four possible combinations.

It was not feasible, under the experimental conditions employed, to control or evaluate precisely the duration of hyperventilation necessary for the production of tetany. Depth and rate of overbreathing were difficult for the experimenter to control and for the subject to regulate, and oftentimes a subject's efforts flagged and the experimenter had to give encouragement and reassurance in order to keep him going. It was assumed that the measurement of alveolar CO_2 pressure at the end of hyperventilation would provide information relevant to the evaluation of differences, between stutterers and nonstutterers, with which the present research was concerned. This appears to be a reasonable assumption since alveolar CO_2 pressure changes slightly, or not at all, after the first five minutes of hyperventilation (*1*), and all subjects in this experiment were hyperventilated at least 10 minutes. For most subjects the duration of hyperventilation was between 15 and 25 minutes and the mean was about 20 minutes.[7]

Subjects. Twenty male college students, receiving speech therapy at the University of Iowa Speech Clinic, and who were considered to be stutterers by themselves and by their clinicians, constituted the experimental group. The age range of this group was 18 to 37 years with a mean age of 22.2 years. Four subjects had been eliminated previously from the experimental group because they had been unable or un-

[5]Chvostek's sign is produced by tapping the facial nerve just below the zygoma. It appears as twitching or spasms of facial muscles or as fine but definite contractions occurring around the eyes, nose, or corners of the mouth. See Best and Taylor (*1*) for further discussion of this and other signs of tetany.

[6]For two of the 20 experimental subjects the between-sessions interval was nine days and for one it was six days.

[7]It is possible to effect some degree of control over duration of hyperventilation by using a tank-type, full-body respirator. See Lillehei and Balke (*22*), for a description of a research project involving this procedure.

TABLE 1. Mean number of total nonfluencies performed by 20 stutterers for three readings under two experimental conditions.

	Reading 1	*Reading 2*	*Reading 3*	*Total*
Hyperventilation	33.1	27.0*	23.4	27.8
No Hyperventilation	33.2	27.0	24.3	28.2
Total	33.2	27.0	23.9	28.0

*Hyperventilation was introduced after Reading 2.

willing to carry through the hyperventilation procedure. The control subjects consisted of 20 male college students who were not and never had been considered stutterers. The mean age of this group was 24.3 years with a range of 20 to 33 years. One subject had been eliminated previously from this group when during a preliminary interview he identified himself as a former stutterer.[8] All experiments were performed in the evening, with the experimenter (Young) as the only observer.

Data Analysis. Five measures of alveolar CO_2 pressure were obtained from each subject: (a) before first reading, (b) after second reading, (c) at beginning of hyperventilation, (d) at end of hyperventilation, and (e) after reading of test passage following hyperventilation. Five additional measures of alveolar CO_2 pressure were obtained from each experimental group subject during the second session, in which a resting period was substituted for the hyperventilation procedure. Two measures of fluency were obtained from the tape recordings of the three readings by the control subjects and the six readings by the experimental subjects. One of these measures was the total number of nonfluencies and the other was rate, defined as the number of seconds required to read the 150-word test passage.[9] The choice of these two measures was based on previous research by Sherman and Trotter (*29*), Sherman, Young, and Gough (*30*) and Young (*36*), which indicated that strong relationships were present between these variables and rated severity of stuttering.

The second experimental session for the stutterers appeared to be necessary since most stutterers demonstrate a decrease in frequency and severity of stuttering over repeated readings of the same material; therefore the effects of hyperventilation and tetany on speech performance would have been confounded with this adaptation effect and could not have been evaluated had not a control condition been included in the design. In addition, there was a need to demonstrate the degree of consistency for alveolar CO_2 pressure on

[8]The use of a third group of subjects had been contemplated in the early planning stages of the present study. This group would have consisted of individuals with speech disturbances such as dysphasia and dysarthria, presumably due to neurological impairment. However, consideration of the physiological effects of hyperventilation, such as hyperpnea, cerebral vasoconstriction and fatigue, persuaded the investigators not to subject persons with the indicated conditions to the experimental procedure.

[9]See Johnson (*18*) for procedures involved in nonfluency analysis.

TABLE 2. Summary of analysis of variance for evaluating the differences among the three readings under two experimental conditions with respect to total number of nonfluencies performed by 20 stutterers.

Source	*df*	*ms*	*F*	*Significance*
Between Subjects	19	4608.68		
Order (C)*	1	5685.63	1.25	NS†
Error b	18	4548.86		
Within Subjects	100	82.55		
Readings (A)	2	891.78	18.32	1%
Conditions (B)	1	3.33	.02	NS
A X B	2	2.01	.15	NS
A X C	2	48.01	.99	NS
B X C	1	300.83	1.43	NS
A X B X C	2	14.11	1.03	NS
Error w	90			
Error w_1	36	48.68		
Error w_2	18	210.82		
Error w_3	36	13.66		
Total	119			

*The two experimental conditions were rotated.
†Not significant at the 5% level.

repeated trials.[10] No comparison of fluency was made between the two groups, since this would not have been relevant to the purpose of the present experiment; comparisons of stutterers and nonstutterers with respect to nonfluency have been presented and discussed elsewhere by Johnson (*14, 16, 17, 18*). The reliability of the experimenter in making the nonfluency analysis was demonstrated in a previous study to be sufficient for the purposes of this investigation (*36*).

Results

Nonfluency Analysis. The nonfluency data obtained from the tape-recorded speech samples of the 20 stutterers and 20 nonstutterers are summarized and evaluated in Tables 1 through 6. The mean total numbers of nonfluencies performed by the 20 stutterers during three readings under two experimental conditions, hyperventilation and no hyperventilation, are to be found in Table 1. Analysis of variance was employed to evaluate the differences between conditions and among readings, and the results of this analysis are presented in Table 2. The obtained significant mean square for Readings (A) and nonsignificant mean squares for Conditions (B) and Readings by Conditions (A $\times$ B) indicate that, for conditions combined, there were systematic differences among the three readings, and that these differences, or trends, when considered for each condition separately, may be assumed to coincide. In other words, it may be inferred that, under the conditions of this experiment, the expected reduction in frequency of nonfluency from reading to reading of the same material is not affected by hyperventilation or hyperventilation tetany.

[10]Data obtained by Sander (*25*) indicated that measures obtained from the nonfluency analysis were highly consistent over a short period of time.

TABLE 3. Mean reading time, in seconds, required by 20 stutterers for three readings of the test passage under two experimental conditions.

	Reading 1	*Reading 2*	*Reading 3*	*Total*
Hyperventilation	144.5	138.0*	124.4	135.6
No Hyperventilation	137.6	136.5	128.1	134.1
Total	141.0	137.2	126.2	134.8

*Hyperventilation was introduced after Reading 2.

TABLE 4. Summary of analysis of variance for evaluating the differences among the three readings under two experimental conditions with respect to reading time, in seconds, required by 20 stutterers.

Source	*df*	*ms*	*F*	*Significance*
Between Subjects	19	62766.26		
Order (C)*	1	75100.02	1.21	NS†
Error b	18	62081.05		
Within Subjects	100			
Readings (A)	2	2369.41	6.07	1%
Conditions (B)	1	70.53	.02	NS
A X B	2	283.61	1.48	NS
A X C	2	830.16	2.12	NS
B X C	1	6780.03	1.93	NS
A X B X C	2	21.06	.11	NS
Error w	90			
Error w_1	36	390.26		
Error w_2	18	3516.80		
Error w_3	36	191.35		
Total	119			

*The two experimental conditions were rotated.
†Not significant at the 5% level.

Essentially the same type of data and analysis are repeated in Tables 3 and 4 as are presented in the two previous tables, but in this instance the measure of fluency was number of seconds required to read the 150-word test passage. The results of this analysis, summarized in Table 4, indicate that the presence of tetany and the other effects of hyperventilation do not alter stutterers' expected speech performance, as far as rate is concerned. These results are not surprising in view of the strong relationship between total number of nonfluencies and reading time.[11]

[11]For these two variables, both Sherman, Young and Gough (*30*) and Young (*36*) found significant correlations of .76.

Analysis of variance could not be used to evaluate the measures of total number of nonfluencies obtained from the tape recordings by 20 nonstutterers, since these measures were not distributed normally, were highly skewed, and had a restricted range. Instead, a Friedman two-way analysis of variance by ranks (*31*) was performed.[12] In the

[12]Scores are ranked for individual subjects with respect to treatments. If there are no differences among the treatments, one would expect the sum of ranks for the three treatments to differ from each other by chance alone. The values for the sum of ranks obtained in the present analysis were 41.0, 36.5, and 42.5 for Readings 1, 2, and 3, respectively. Chi square, df = a-1, is computed as follows:

$$\chi^2_r = 12\ R_j^2 / Na(a+1) - 3N(a+1)$$

where R_j = sum of ranks, N = number of subjects, and a = number of treatments.

TABLE 5. Summary of analysis of variance for evaluating the differences among the three readings with respect to reading time, in seconds, required by 20 nonstutterers.

Source	*df*	*ms*	*F*	*Significance*
Readings (A)	2	23.62	4.64	5%
Subjects (S)	19	118.59		
A X S	38	5.09		
Total	59			

TABLE 6. Mean values of alveolar air carbon dioxide tension for 20 stutterers and 20 nonstutterers under five experimental conditions.*

	Conditions					
	1	*2*	*3*	*4*	*5*	*Total*
Stutterers	40.2†	40.1	39.2	21.2	30.9	34.3
Nonstutterers	40.4	40.5	39.1	22.5	32.2	35.0
Total	40.3	40.3	39.2	21.8	31.6	34.7

*1: before reading; 2: after reading; 3: at beginning of hyperventilation; 4: at end of hyperventilation; 5: after reading following hyperventilation.

†Values in terms of mm Hg.

present study, the value computed was .98 ($df = 2$), which was not significant at the 5% level.

The mean values of reading time required by the stutterers for the three readings of the test passage under the hyperventilation condition were 59.2, 57.8, and 60.0 seconds, respectively. It can be seen that these means for the three readings show the same trend that the sum of ranks did in the test mentioned just previously; that is, there was an increase in fluency, defined as rate, from the first to the second reading, and then a decrease in fluency following hyperventilation. Analysis of variance, the results of which are presented in Table 5, was used to evaluate the differences among the means. The mean square for Readings (A) was significant at the 5% level, indicating that there were significant differences among the three readings in terms of reading time. The critical difference[13] necessary for significance at the 5% level is 1.44. Only the difference between Reading 2 and Reading 3 reached significance at the 5% level. The implications which may be drawn from these statistical tests, involving total number of nonfluencies and reading time, are that hyperventilation and tetany are followed by slight decrease in the speech fluency of the nonstutterers, insofar as these measures are to be considered as indicators of fluency.

Carbon Dioxide Tension. Alveolar air carbon dioxide tension, in terms of mm Hg, was measured under five experimental conditions for the 20 stutterers and 20 nonstutterers, and the means for these measures are to be found in Table 6. From an examination

[13]Critical difference $= t_{.05}\,(2ms_{AS}/s)^{1/2}$.

TABLE 7. Summary of analysis of variance for evaluating the differences between stutterers and nonstutterers with respect to alveolar air carbon dioxide tension under five experimental conditions.

Source	*df*	*ms*	*F*	*Significance*
Between Subjects	39	54.01		
Groups (B)	1	19.22	.35	NS*
Error b	38	54.93		
Within Subjects	160	69.24		
Conditions (A)	4	2580.52	529.92	1%
A X B	4	4.02	.83	NS
Error w	152	4.87		
Total	199			

*Not significant at the 5% level.

of this table it can be seen that at the end of hyperventilation, Condition 4, the mean value of alveolar CO_2 pressure is almost half the normal or resting value in Condition 1, and also that in Condition 5, after reading following hyperventilation, alveolar CO_2 pressure is only about three-fourths of the normal value. This would imply that the effects of hyperventilation were still partially present during the third reading of the test passage, a circumstance quite important to the design of this study. Analysis of variance, as shown in Table 7, was employed to evaluate the differences between stutterers and nonstutterers with respect to the variable under consideration. Observation of the data reported in Table 6 alone would indicate that there were no significant differences between the two groups and that no interaction was present between the five conditions and the two groups. The statistical analysis in Table 7 supports this observation, since the mean square for Groups (B) and the mean square for interaction (A $\times$ B) were not significant at the 5% level. It seems unlikely that a Type II error has occurred, that is, that a false null hypothesis has been accepted, in view of the small differences between the means of the two groups for each condition. The significant mean square for Conditions (A) was to be expected, since this variable was under control of the experimenter, and, in fact, was part of the experimental design.

Data were available to evaluate the consistency of alveolar CO_2 pressure measures for the stutterers since a second experimental session was added for this group. Although alveolar CO_2 pressure was measured five times in the second or control session, only the first

TABLE 8. Mean values of alveolar air carbon dioxide tension for 20 stutterers before and after reading under two experimental conditions.

	Before Reading 1	*After Reading 2*	*Total*
Hyperventilation	40.2*	40.1	40.2
No Hyperventilation	41.1	41.6	41.3
Total	40.6	40.9	40.7

*Values in terms of mm Hg.

TABLE 9. Summary of analysis of variance for evaluating the differences between alveolar air carbon dioxide tension for 20 stutterers for two readings under two experimental conditions.

Source	*df*	*ms*	*F*	*Significance*
Between Subjects	19	33.23		
Order (C)*	1	5.89	.17	NS†
Error b	18	34.75		
Within Subjects	60	6.73		
Readings (A)	1	.90	.13	NS
Conditions (B)	1	28.44	3.94	NS
A X B	1	2.21	.59	NS
A X C	1	22.37	3.24	NS
B X C	1	.02	.00	NS
A X B X C	1	28.20	7.49	5%
Error w	54	5.96		
Error w_1	18	6.89		
Error w_2	18	7.22		
Error w_3	18	3.76		
Total	79			

*The two experimental conditions were rotated.
†Not significant at the 5% level.

two measures obtained are pertinent to the problem of alveolar CO_2 pressure consistency. Mean values of alveolar CO_2 pressure before and after reading under the two conditions are presented in Table 8. There was an interval of seven days for almost all subjects between the two sessions, and the order of the sessions was rotated. Table 9, a summary of the analysis of variance employed to evaluate these means, indicates no significant differences between measures before and after reading for conditions combined or between the two experimental sessions for readings combined. This apparent consistency is reassuring in view of the analysis of variance presented in Table 7 which indicated that no differences were present in alveolar CO_2 pressure between the stutterers and the nonstutterers.

Discussion

The use of a reading passage, choice of nonfluency measures, and inclusion of stutterers representing various levels of severity, were dictated by the experimental design, previous research, and other practical considerations. For these reasons and others to be indicated certain limitations are to be placed on generalizations drawn from the obtained data. First, stutterers and nonstutterers were hyperventilated only to the point at which the experimenter considered Chvostek's sign to be positive or carpopedal spasms to be present. Second, there are many other physiological consequences of hyperventilation which may have been confounded with the effects of tetany, both for the stutterers and the nonstutterers. Also to be considered is the fact that the subjects, sitting on a hospital litter, read a relatively short and emotionally neutral prose passage, hardly a usual type of speaking situation. However, of the stutterers' 120 readings, roughly one-third were considered to contain fewer than 10 nonfluencies, and it is to be

assumed that if deterioration in fluency occurs it is to be observed more easily in the speech of mild than of severe stutterers since the resulting difference would be more conspicuous.

While generalization is to be judiciously limited, it is to be duly considered that if stutterers are supposed to be precariously poised on the brink or threshold of some imbalance, biochemical or neurological, it might reasonably be expected that hyperventilation and its several consequences—such as tetany, hyperpnea, apnea, respiratory alkalosis, and increased irritability of muscle tissue—would be followed by some more or less readily detectable change in the fluency of their speech. The data obtained indicate quite clearly that this was not the case. It might be postulated that stutterers have an excess of carbon dioxide, hence the decreased fluency for the nonstutterers after hyperventilation and no change for the stutterers, but the analysis indicates that there are no differences with respect to alveolar CO_2 pressure under the five experimental conditions examined. Considering the fluctuation of alveolar CO_2 pressure over time, and the experimental error possible in measuring alveolar CO_2 pressure, it is surprising that the differences between two groups of subjects would be as small as those obtained in this study. The essential similarity of the two groups in this respect provides no support for the hypothesis that in stutterers tetany is latent and readily activated by comparatively slight modification in CO_2 tension and other relevant conditions.

In general, the findings of this investigation do not lend support to the general theory that stuttering, considered as nonfluency of the speech of persons classified as stutterers, is due to organic factors. It does not follow, however, that speech fluency, *per se*, may not be functionally related to organic factors; indeed, in the present study hyperventilation and induced tetany were followed by a slight decrease in the speech fluency of the nonstutterers. A major implication of the data is that the problem generally referred to as stuttering is to be differentiated from the problem of speech nonfluency, as such. In light of this implication, the present study may be viewed most appropriately as an investigation of the relationship between neuromuscular and biochemical variation and change in the fluency and rate of speech—in speakers classified either as stutterers or as nonstutterers.

Summary

Data were obtained to test the hypothesis that stuttering is associated with latent tetany. Twenty stutterers and 20 nonstutterers read a 150-word passage three times, and were hyperventilated between the second and third readings until signs of tetany appeared. It was concluded that the stutterers and nonstutterers did not differ significantly with respect to alveolar air carbon dioxide tension, and that there were no significant changes in the speech fluency of the stutterers as a result of hyperventilation and tetany. There was, however, a slight decrease in the speech fluency of the nonstutterers following induction of tetany through hyperventilation. The findings imply that the problem called stuttering is to be differentiated from the problem of speech nonfluency, as such.

Acknowledgment

The authors wish to acknowledge the assistance of Mr. Newton Whitman, then a graduate student in the Department of Chemistry, University of Iowa, in the collection of preliminary data, and of Dr. Paul Seebohm, Professor of Internal Medicine, University of Iowa, who, as codirector with Dr. Bedell of the Pulmonary Research Laboratory, authorized essential arrangements for collection of data.

References

1. Best, C. H., and Taylor, N. B., *Physiological Basis of Medical Practice*. Baltimore: Williams and Wilkins, 1955.
2. Bloodstein, O., Stuttering as an anticipatory struggle reaction. In J. Eisenson (Ed.), *Stuttering: A Symposium*. New York: Harper, 1958.
3. Brown, E. B., Physiological effects of hyperventilation. *Physiol. Rev.*, 33, 1953, 445-471.
4. Collier, C. R., Affeldt, J. E., and Farr, A. F., Continuous rapid infrared CO_2 analysis. *J. Lab. clin. Med.*, 45, 1955, 526-539.
5. Collip, J. B., and Backus, P. L., The effect of prolonged hyperpnea on the CO_2 combining power of the plasma, the CO_2 tension of the alveolar air and the excretion of acid and basic phosphate and ammonia by the kidney. *Amer. J. Physiol.*, 51, 1920, 568-579.
6. Eisenson, J., A perseverative theory of stuttering. In J. Eisenson (Ed.), *Stuttering: A Symposium*. New York: Harper, 1958.
7. Gellhorn, E., and Kraines, S. H., Word association as affected by deficient oxygen, excess of CO_2 and hyperpnea. *Arch. Neurol. Psychiat.*, 38, 1937, 491.
8. Gellhorn, E., and Speisman, I. G., Influence of hyperpnea and of variations of O_2 and CO_2 tension in the inspired air upon hearing. *Amer. J. Physiol.*, 112, 1935, 519.
9. Gellhorn, E., and Speisman, I. G., Influence of hyperpnea and of variations of O_2 and CO_2 tension in the inspired air upon nystagmus. *Amer. J. Psysiol.*, 112, 1935, 622.
10. Gellhorn, E., and Speisman, I. G., Influence of hyperpnea and of variations of O_2 and CO_2 tension in the inspired air upon after-images. *Amer. J. Physiol.*, 112, 1935, 620.
11. Glauber, I. P., The psychoanalysis of stuttering. In J. Eisenson, (Ed.), *Stuttering: A Symposium*. New York: Harper, 1958.
12. Grant, S. B., and Goldman, A., A study of forced respiration: Experimental production of tetany. *Amer. J. Physiol.*, 52, 1920, 209-232.
13. Hill, H., Stuttering: I. A critical review and evaluation of biochemical investigations. *J. Speech Dis.*, 9, 1944, 245-261.
14. Johnson, W., (Ed.), assisted by R. Leutenegger, *Stuttering in Children and Adults*. Minneapolis: Univ. of Minnesota Press, 1955.
15. Johnson, W., Perceptual and evaluative factors in stuttering. In L. E. Travis (Ed.), *Handbook of Speech Pathology*. New York: Appleton-Century-Crofts, 1957. Reprinted, *Folia phoniatr.*, 8, 1956, 211-233.
16. Johnson, W., Brown, S. F., Curtis, J. F., Edney, C. W., and Keaster, Jacqueline, *Speech Handicapped School Children* (rev. ed.). New York: Harper, 1956.
17. Johnson, W., and associates, *The Onset of Stuttering: Research Findings and Implications*. Minneapolis: Univ. Minnesota Press, 1959.
18. Johnson, W., and associates, Speech fluency norms. Unpublished report of research carried out at University of Iowa under grant from Louis W. and Maud Hill Family Foundation, St. Paul, Minn.
19. Johnson, W., Stearns, Genevieve, and Warweg, Edna, Chemical factors and the stuttering spasm. *Quart. J. Speech*, 19, 1933, 409-415.
20. Kopp, G. A., Metabolic studies of stutterers. I. Biochemical study of blood composition. *Speech Monogr.*, 1, 1934, 117-132.
21. Lewis, B. I., Chronic hyperventilation syndrome. *J. Amer. med. Ass.*, 155, 1954, 1204-1208.
22. Lillehei, J. P., and Balke, B., Studies of hyperventilation. USAF School of Aviation Medicine, Randolph Field, Texas, 1955.
23. Rahn, H., Otis, A. B., Hodge, M., Epstein, M. A., Hunter, S. W., and Fenn, W. O., The effects of hypocapnia on performance. *J. Aviat. Med.*, 17, 1946, 164.

24. Rushmer, R. F., and Bond, D. O., The hyperventilation syndrome. *War Med.*, 5, 1944, 302-303.
25. Sander, E. K., Temporal reliability (24 hours) of the reading and speaking tasks of the Iowa Speech Fluency Test administered to 40 stutterers. M.A. thesis, Univ. Iowa, 1959.
26. Scow, J., Krasno, L. R., and Ivy, A. C., Immediate and accumulative effect on psychomotor performance of exposure to hypoxia, high altitude and hyperventilation. *J. Aviat. Med.*, 21, 1950, 79.
27. Shackson, R., An action current study of muscle contraction latency with special reference to latent tetany in stutterers. *Arch. Speech*, 1, 1936, 87-111.
28. Sheehan, J., Conflict theory of stuttering. In J. Eisenson (Ed.), *Stuttering: A Symposium*. New York: Harper, 1958.
29. Sherman, Dorothy, and Trotter, W. D., Correlation between two measures of the severity of stuttering. *J. Speech Hearing Dis.*, 21, 1956, 426-429.
30. Sherman, Dorothy, Young, M. A., and Gough, K. H., Comparison of three measures of stuttering severity. *Proceedings of the Iowa Academy of Science*, 65, 1958, 381-384.
31. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1956.
32. Travis, L. E., The unspeakable feelings of people with special reference to stuttering. In L. E. Travis (Ed.), *Handbook of Speech Pathology*. New York: Appleton-Century-Crofts, 1957.
33. Van Riper, C., *Speech Correction Principles and Methods* (3rd ed.). New York: Prentice-Hall, 1954.
34. West, R., An agnostic's speculations about stuttering. In J. Eisenson (Ed.), *Stuttering: A Symposium*. New York: Harper, 1958.
35. Wischner, G. J., Stuttering behavior and learning: a preliminary theoretical formulation. *J. Speech Hearing Dis.*, 15, 1950, 324-335.
36. Young, M. A., Correlational analysis of speech nonfluencies and rated severity of stuttering. M.A. thesis, Univ. Iowa, 1958.

The Latin Square Design in Speech and Hearing Research

LEONARD S. FELDT

Even a casual review of the speech and hearing research literature for the past 10 years reveals quite remarkable progress in the improvement of the typical experimental design. Within this period experimenters have advanced from almost total reliance (often inappropriate reliance) on *t* tests, chi square tests, and simple product-moment correlation coefficients to sophisticated use of complex analysis of variance designs. The adoption by one discipline of statistical techniques originally developed by another is often accompanied by a high proportion of misapplications. Many flagrant examples of such misuse of the methods of analysis of variance can be found in the psychological literature of 15 or 20 years ago, when the techniques were newly borrowed from agricultural research. The transition of these same techniques from experimental psychology to speech and hearing research has undeniably involved numerous instances of this unfortunate trend. Yet it may be fairly said that as the body of trained research workers has grown, the proportion of poorly conceived and inappropriately executed analyses has declined to the point of favorable comparison with research in the other behavioral sciences.

The most frequent errors occurring today involve the more complex analyses, particularly those designs involving several experimental dimensions and repeated measurements of the experimental subjects. The Latin square design has been especially subject to abuse. The purpose of this paper is to clarify the nature of this design, to discuss the assumptions which underlie it, and to outline some of its advantages and limitations in speech and hearing research.

As employed in the behavioral sciences, the basic Latin square design represents a variation on the two-dimensional factorial analysis. Both of the experimental variables are 'assignable'; that is, the various levels of each treatment represent conditions which may be manipulated by the experimenter. They are imposed at will on any specified subject rather than accepted as an unalterable characteristic of the subject. The important differences between the Latin square and the factorial design are that (a) the same numbers of levels must be represented in all experimental variables and (b) it must be administratively and experimentally feasible to expose an experimental subject to more than one level

Leonard S. Feldt (Ph.D., University of Iowa, 1954) is Associate Professor of Education, University of Iowa, and Statistical Consultant, *Journal of Speech and Hearing Research.*

of each treatment. In an A × B factorial design, with a levels of both variables, there are a total of $a \cdot a$ or a^2 treatment combinations. In the Latin square design, each subject is required to take a of these combinations, among which each level of each variable occurs once and only once. Thus, if the treatments are designated A_1, A_2, A_3, B_1, B_2, and B_3, Subject 1 might be assigned these combinations: A_1 with B_2, A_2 with B_3, and A_3 with B_1. Subject 2 might be assigned A_1 with B_3, A_2 with B_2, A_3 with B_1. In each case every level of A and every level of B are represented, though the combinations vary and no subject is exposed to any level more than once.

Many experimenters may not realize that this type of design represents a considerable modification of the Latin square from its original conception in agricultural research. In the latter field the design involved three, not two, experimental variables and the nature of the variables had a marked effect on the assumptions underlying the design. Two specific examples might clarify the differences. Suppose an experimenter wishes to study the effects of delayed sidetone on some aspect of speech, such as number of nonfluencies or speech blocks, under varying conditions of background noise. Three levels of noise (factor A) are to be employed at each of three levels of delay (factor B). Knowing the Latin square design would improve experimental precision (a topic discussed in a later section), the experimenter would randomly subdivide his subjects into three equal groups (I, II, and III) and assign the groups to treatment combinations as indicated here.

	Noise 1	Noise 2	Noise 3
Delay 1	Gp. I	Gp. II	Gp. III
Delay 2	Gp. III	Gp. I	Gp. II
Delay 3	Gp. II	Gp. III	Gp. I

Consider now a typical agricultural experiment involving the effect of various dosages of a new drug (factor C) on the average birthweight of litters of pigs. The experimental animals are to be drawn from three randomly selected herds (factor A), and the animals are to farrow within three randomly selected intervals of the year (factor B). The experiment might be diagrammed as follows:

	Herd 1	Herd 2	Herd 3
Time 1	Dose 1	Dose 2	Dose 3
Time 2	Dose 3	Dose 1	Dose 2
Time 3	Dose 2	Dose 3	Dose 1

Several important differences are brought out by these examples. First, the intrinsic nature of factors A and B varies from one experiment to the other. In the speech experiment these factors represent the important effects under study. The differences between groups are entirely random, or would have been had all groups been given the same treatments. In the agricultural experiment, 'Herds' and 'Times' are factors of little interest, except to the extent that they make possible greater generality and possibly greater precision in the evaluation of factor C. Thus

the random or nonrandom nature of the variables in these two areas is often reversed. A second important difference lies in the manner in which the levels of each variable are defined. In the speech experiment, the levels are chosen only after careful consideration of the overall experimental situation, and conclusions will pertain only to the specific levels selected. In the agricultural experiment, the three 'levels' of A and B are randomly chosen from large populations of such levels, and conclusions will apply to the entire populations. A third crucial difference concerns the character of the interaction of factors A and B. In the speech experiment this effect might well be as important as (or more important than) the two main effects. The experimenter would exercise considerable care to keep this effect uncontaminated by factors which could systematically affect all of the subjects under any combination of treatments. The presence of random differences between groups would not invalidate the test of this effect, since such differences would be fully reflected in the error term used to assess significance. In the agricultural experiment, the interaction between Herds and Times might be regarded, were it not for the presence of factor C, as reflecting only sampling and experimental error. The main effect of C, computed by combining the cells involving each specific dosage, would therefore be uncontaminated by any real interaction between factors A and B. The variance analysis would include mean squares only for the three main effects and residual. Since no intrinsic interactions are presumed to exist, no mean square is computed for them.

It was a number of years before the majority of experimenters in psychology and education grasped these important differences. Many a psychological experiment followed the agricultural model faithfully, even though neither factor A nor B could be considered random nor the interaction of A and B reasonably expected to involve only error variance. The effect of variable C was duly evaluated by combining the cells into three sets of three each: $A_1B_1C_1 + A_2B_2C_1 + A_3B_3C_1$, $A_2B_1C_2 + A_3B_2C_2 + A_1B_3C_2$, and $A_3B_1C_3 + A_2B_3C_3 + A_1B_2C_3$. The differences between the means of these sets did present the main effect of factor C, of course, but as sophisticated experimenters now recognize, they also partially reflected the A $\times$ B interaction. Thus, as McNemar (*5*) has pointed out, such an experiment could never adequately assess the importance of factor C, the crucial experimental variable.

Research workers in speech pathology and audiology have profited from these early errors, perhaps because many of today's workers have been trained by a generation of psychologists that first committed, then recognized, them. But other subtleties of the Latin square are still often overlooked, particularly those which concern the interaction of the experimental variables and the assumptions underlying the statistical tests. These are discussed in the following sections.

Interaction in the Latin Square Design

Gourlay (*3*) has made a very worthwhile distinction between two types of interaction which may arise in the Latin square design. The first is that intrinsic

interaction which exists between at least two levels of factor A and two levels of factor B and which would be revealed in an orthodox factorial design of sufficient precision. It is the type of interaction which would exist within independent treatment populations, and constitutes the interaction of prime interest to the experimenter. The second type is no less real but considerably less interesting. It is the interaction which arises solely as a result of the repeated use of experimental subjects—the interaction which would occur in related treatment populations of a Latin square design, but would not occur in independent treatment populations of an orthodox factorial design. In discussing designs involving replicated measurement, Lindquist (*4*) implicitly makes the same distinction when he emphasizes the importance of the assumption that the treatments must have little residual effect on the subject. Unless the cumulative effect of treatments is systematically controlled or is of primary interest, each experimental condition must 'wear off' and leave the subject free to react quite naïvely to any subsequent treatment he may receive. Should the first treatment combination influence the second, and the first plus the second influence the third, the validity of the experiment would clearly be impaired.

In many areas of speech research the presumption of no residual effects seems quite defensible. Many experimental conditions involve no learning, no practice effects, no fatigue factors, and no interference or inhibition phenomena. A subject's perception of pure tones under various conditions, for example, can be measured on occasions sufficiently separated in time to render one listening session of negligible influence on the next. Speech samples of stutterers can be collected under a variety of conditions that would result in little 'intermingling' of treatments. Some experiments involve data obtained from subjects who are quite unaware that they are participating in a controlled study, and hence the treatment conditions affect one another no more than one normal life activity affects another. But in the majority of experiments the influence of one treatment on another cannot be assumed negligible, and the experimenter must employ a design which exercises some measure of control over cumulative effects. To ignore them is to fail to recognize that the observed interaction may not be a true treatment interaction but rather a carry-over effect of a highly artificial kind. Even the absence of interaction cannot be unambiguously interpreted, since the residual effect might cancel or offset the more important component of the interaction.

Residual or carry-over effects may actually be of two kinds, one somewhat simpler than the other. Treatment A_1 may provide beneficial practice for treatment A_2 and these in turn may contribute to even better performance on A_3. If the sequence were changed to A_2–A_3–A_1, then A_1 would stand to benefit to precisely the same degree. Such an effect might be called a rank-order carry-over—an advantage (or disadvantage) accruing to the second, third, . . . treatments as a direct result of practice, learning, fatigue, motivational changes, and the like. The effect is simply additive, a function of 'firstness,' 'secondness,' 'thirdness,'. . . . The

second carry-over effect is somewhat subtler. It arises whenever the practice, adaptation, or fatigue factors generated by one treatment condition are greater than those produced by another. Perhaps condition A_1 provides better practice for A_2 than A_3 does for A_2. Thus, the combination of treatment A_1 followed by A_2 followed by A_3 may yield, on the average, higher (or lower) scores than the combination treatment A_3 followed by A_2 followed by A_1. In each case all treatments and all positions (first, second, and third) are represented, yet the combinations are not equivalent. In such a situation, sequence of treatments makes a difference, and it seems reasonable to call this phenomenon a sequence carry-over.

Either, both, or neither of these carry-over effects can potentially operate in any Latin square. It thus becomes highly crucial that the experimenter, unless he has good reason to expect that they won't arise, plan the design to cope with this eventuality. This may be done in three ways: (a) list all possible sequences of treatments and choose one randomly and independently for each subject; (b) employ a counterbalancing arrangement which permits each treatment to occur equally often in each administrative position; or (c) pick out a consistent, prescribed order for the A treatments to apply to all subjects and let the orders occur as they may so far as B is concerned. Each of these options has both advantages and disadvantages, but the counterbalancing approach is probably the most generally useful solution to the problem. Its only drawback, which the conscientious experimenter can resolve, is the complexity it adds to the analysis and reporting of the experimental data.

Counterbalancing Order in the Latin Square

The counterbalancing solution is easily stated, less easily grasped. In essence, it dictates that rank-order position (first, second, . . .) must itself become a treatment variable. The complexity thus introduced is that contributed by an extra variable. The one-dimensional experiment becomes two-dimensional; the two-dimensional experiment becomes three-dimensional. In addition, some seemingly troublesome interactions may be called forth. As it happens, these interactions, even when significant, can be dealt with relatively easily.

The simplest Latin square in which 'order' represents a dimension in the design involves only one experimental variable. A simpler version of the sidetone experiment described earlier will serve as an example. The auditory feedback, at one specified level of delay, is to be presented against two levels of background noise. Each subject is to be exposed to both levels of noise; but the sequence to be followed is Loud-Soft for one half of the subjects, Soft-Loud for the other half. The measure is the number of speech blocks. The design could be represented as follows:

	Noise Level 1	Noise Level 2
First Treatment	A_1O_1 Group I	A_2O_1 Group II
Second Treatment	A_1O_2 Group II	A_2O_2 Group I

Had the experimenter a greater number of subjects, he might perform a

sufficiently precise experiment with an independent group at each noise level and make a simple *t* test to assess the difference between the two conditions. A more sophisticated investigator would recognize, however, that administration of both conditions to each subject effectively doubles the number of observations and makes the comparison of A_1 against A_2 a comparison based on related measures. To the extent of the relationship, such a design would permit a more precise evaluation of treatment differences. But the sophisticated experimenter is also aware of possible carry-over effects. Perhaps there is an adaptation effect that lessens the disruptive influence of delayed sidetone with repeated presentations (rank-order carry-over), that is, the number of speech blocks decreases with repeated presentations. Perhaps, also, the degree of adaptation under the sequence 'loud background followed by soft background' differs from that under 'soft background followed by loud background' (sequence carry-over). In view of these possibilities, order is introduced as a second factor in the design, and it now becomes a Latin square rather than a simple comparison involving a *t* test. Group I, a random half of the available subjects, takes A_1 (loud background) followed by A_2 (soft background); Group II takes the conditions in the reverse order.

The top row of the cells in the block diagram includes those measures obtained under the first treatment received by each group; the bottom row of cells includes measures taken under the second treatment. Thus, the rank-order carry-over about which the experimenter is concerned becomes a main effect in the analysis. It is of little interest and is probably not tested; but it is isolated from effects which *are* of interest, and it is prevented from influencing the comparison of A_1 with A_2 in either a biasing or random fashion.

The interaction between order and treatments is evaluated by the comparison of the mean of cells A_1O_1 and A_2O_2 combined against the mean of cells A_2O_1 and A_1O_2 combined. Stated more simply, sequence A_1–A_2 is compared to sequence A_2–A_1. Both treatments and both rank-order positions are represented, hence a difference can reflect only the second of the potential carry-over effects, that is, the effect due to sequence differences.

Since both a first and a second administration are represented under each noise level, the presence of a rank-order carry-over is of little consequence. If the adaptation after one session equals five fewer speech blocks, the second set of measures under each noise level will be reduced by this amount. The presence of a significant interaction is in some ways more costly, but it in no way invalidates the experiment. This outcome merely leads the experimenter to the conclusion that there was no unbiased way in which his subjects could be used twice and still duplicate the treatment differences which an independent-group comparison would have revealed. The conclusion to be drawn is that comparisons of treatment effects, in all positions but the first, are not to be trusted. The experimenter has lost his gamble to gain precision and must resort to the *t* test or *F* test which involves only the means for treatments in the first rank-order position. But it has been, or usually is, a good gamble. The recourse to an unbiased compari-

son still exists in the top row of means. The only loss occurs in the outside chance that had a simple-randomized design been planned from the outset, more subjects might have been employed. However, in most experimental studies the investigator will generally employ all of the subjects he can conscript or afford no matter what design is to be employed. Hence even this loss may be more apparent than real.

The general conclusions reached in this simple experiment apply to studies of larger scope. The main effect of order—significant or not—can generally be ignored unless, of course, the experimenter has some interest in it. Interaction will be tested for assessment of the artificial sequence carry-over. Significance will lead to the examination of treatment effects as reflected in the first administration of each condition.

Often the experimental conditions may make it quite unwise to repeat precisely the same criterion task when the subjects are exposed to a second, third, or fourth treatment. Stutterers, for example, are known to exhibit fewer blocks on rereading a specified test passage. A filmed test of lipreading undoubtedly leaves an imprint on the subject which would not be erased were he exposed to the film a second time. In such instances, parallel forms of the criterion task are called for, usually one for each treatment the subject receives. Such parallel forms can be effectively incorporated into the Latin square as a treatment variable quite analogous to rank-order discussed above. Small or moderate differences in the difficulty of only roughly equated tests are thus fully controlled, since each form, like each order, will be equally represented under all treatment conditions.

A real interaction of treatments with test forms represents a more serious difficulty, however. Such an interaction reflects fundamental differences between forms—a content uniqueness which is brought out by the distinct elements in at least one pair of treatments. If such is revealed, the experimenter might well consider postponing further investigation of his treatments until the most valid of his forms can be identified. Happily, such interactions are relatively rare.

When the Latin square includes three or more levels of the treatment variable, and order is counterbalanced, the sequence carry-over (order by treatment interaction) is in part a 'between-subjects' effect. This component is quantified by the mean square for groups. Some experimenters have interpreted this mean square as a measure of the 'order effect' and treated it as a main effect of the experiment. This is a natural interpretation, perhaps, since 'sequence' and 'order' are used almost interchangeably in common speech. However, in the Latin square design, order and sequence mean quite different things. To define group differences as order differences is to give a new and confusing definition to the concept of order. The interaction of this 'order' variable with treatments includes a component which reflects precisely the same phenomenon that contributes to the main effect itself (sequence differences). The interaction also includes a component reflecting rank-order differences. Clearly, the definition of group differences as 'order effects' may engender considerable confusion in the

interpretation of the tests of significance. Experimenters who wish to counterbalance order should have its definition clearly in mind before setting up the analysis.

Perhaps a more common experimental error is to ignore entirely the effects of counterbalancing. Many experimenters recognize intuitively the advantages of counterbalancing test forms or order and adopt such procedures in the administration of the experiment. Intuition may not impel the experimenter to modify the analysis accordingly, however. In such a case the order and form effects contribute heavily to the error variance, but do not contribute at all to the treatments effect. The net result is a marked increase in the probability of retaining a false null hypothesis. Whenever such controls are instituted in an experiment, a variance component for the controlled effect and its interactions should always be extracted in the analysis of variance. Failure to extract these components can be seriously damaging to the precision of the experiment.

Confounding Test Forms and Rank Order

Not infrequently, situations arise in which the effects of rank-order carry-over must be held constant, sequential carry-over must be investigated, and parallel forms must be employed. For example, in the delayed sidetone experiment described earlier, the investigator might deem it wise to utilize a parallel reading passage rather than the same reading passage when subjects are exposed to the second level of background noise. In such a case it would appear that a third dimension is called for—a Latin square involving two levels of treatments, orders, and forms. Many experimenters have used just such a design. However, one of the subtle aspects of Latin squares which involve rank-order is that the forms effect can be simultaneously accommodated, along with order, within the framework of a two-dimensional design. This is accomplished by the simple expedient of purposely confounding forms with orders. In the delayed sidetone example, Passage I would be consistently used with the first treatment received by each subject, be it the Loud or Soft noise level, while Passage II would be consistently used with the second. The row effect would now become one of 'Form plus Administrative Order.' Ordinarily, of course, confounding is both dangerous and inadvisable. But since neither of these main effects holds any intrinsic interest for the experimenter, their net effect is also of no consequence.

The interaction of such a combined factor with treatments would also become a combined effect involving the interaction forms by treatments, forms by orders, and forms by orders by treatments, as well as the sequence carry-over effect. However, in experiments where the forms clearly involve similar content or demand identical skills, interactions with forms are almost certain to be very small or negligible. Thus, a significant interaction can almost always be attributed to a sequence carry-over. Whatever its origin, an interaction will generally lead to examination of the still valid simple effect based on the first administration of each treatment. As insurance against such an eventuality, the experimenter might

well use his 'best' form, if he can designate one, for the first administration of each treatment. Then if a simple effects test is necessary, it will be made under optimal conditions.

The confounding of forms with administrative orders introduces welcome simplifications in both the analysis and reporting of the experiment, although the legitimacy of the procedure will generally have to be defended at least briefly in the research report. In addition, some increase in the precision of the experiment will be secured. Were forms and rank-order treated separately, estimates of the following effects would be isolated: orders, forms, orders by treatments, forms by treatments, forms by orders, and forms by orders by treatments. If the number of conditions exceeds two, each of these effects would be a 'within' factor or would have a 'within' component, and their degrees of freedom would total $(a-1) \times (a^2-2)$. With forms and rank-order confounded, only the one main effect and one interaction of this combined factor with treatments need be extracted, at a cost of only $(a-1)\ (a-1)$ degrees of freedom. Since the degrees of freedom of the unwanted terms must be deducted from those for the error term and since the population value of the error variance is the same under both approaches, the confounding technique, with its higher *df* for error, is more precise. In a three-level experiment this difference would be 10 degrees of freedom; in a four-level experiment, 33 degrees of freedom. No loss of valuable information is suffered by the confounding and the recourse to a valid test of simple effects remains available, if needed.

Improvements in Precision Effected by Latin Square Designs

The precision of an experimental design may be evaluated in a number of ways. The size of the population error variance, the power of rejecting false hypotheses, and the degrees of freedom available for error are three indexes often taken to reflect precision. The most generally satisfactory index for comparing various designs, however, is the average size of the standard errors of the differences for all pairs of treatment means. (With different numbers of observations, the standard error would vary from pair to pair.) When the degrees of freedom for error are approximately equal for two designs, the one yielding the smaller average standard error will be the more precise. This index provides a simple basis for evaluating the improvement in precision which the Latin square attains over the independent-groups design which might otherwise be employed.

If the Latin square involves order, forms, or the two confounded, it is appropriately compared to the alternative of a simple-randomized design. Let σ_e^2 represent the common within-treatments variance, N the total number of subjects available, *a* the number of treatments, and n, that is N/a, the number of subjects per treatment in the simple-randomized experiment. The average standard error of differences in this analysis is $(2\sigma_e^2/n)^{1/2}$ and it is estimated by $(2ms_w/n)^{1/2}$. In the Latin square design the estimator of the average standard error of differences, $(2ms_{error\ (w)}/N)^{1/2}$, is probably more familiar to researchers than the population quantity it estimates. It may be shown (2) that the mean square for

error (w) in a Latin square is an estimate of $\sigma_e^2 (1 - \rho)$, where ρ is the average correlation between all pairs of related measures. (In a three-level experiment, three such correlations could be computed for each of the three groups; $\bar{\rho}$ represents the average of the nine population coefficients.) Thus, the average standard error of differences would be $(2\sigma_e^2[1 - \rho]/N)^{1/2}$, the N in the denominator reflecting the fact that each treatment mean is based on all of the experimental subjects.

The ratio of the standard error of the Latin square to that of the simple-randomized design is $([1 - \bar{\rho}]/a)^{1/2}$. Since the value of $\bar{\rho}$ in speech experiments almost always falls between .40 and .80 and since the most common experimental designs include two to four treatment conditions, this ratio will generally fall between .22 and .55. This represents a rather startling reduction in the magnitude of the standard error, roughly 50 to 75%. Translated into power terms, such a reduction would mean that if a two-treatment, simple-randomized experiment had 50 chances in 100 of revealing a true difference, the Latin square would have better than 94 chances in 100. If the simple randomized design had 80 chances in 100 of revealing the difference, the Latin square would have better than 99 chances in 100. Clearly, the Latin square holds great promise in increasing the precision of experimental comparisons.

Often two or more experimental dimensions are included in the Latin square. For example, the experiment outlined earlier could include several delay intervals as well as several levels of background noise. For simplicity's sake the experimenter might prefer to limit the design to two dimensions and not include order as a third. In that case the treatment sequence should be independently determined for each subject. If the basic assumptions discussed later are met, a valid test of the two main effects and their interaction could be made against the mean square for error (w). Such a test is always valid regardless of the existence of rank-order and sequence carry-over effects. Independent randomization, subject by subject, insures that these effects will contribute appropriately to treatments effects, interaction, and error. To the extent, however, that such effects are greater than zero they reduce the precision of the experiment. Therefore, if they may be anticipated before the experiment is begun, it would surely be wise to include order as a third dimension.

Because randomization always provides a valid test, an interesting question arises. Suppose order has been counterbalanced and the experiment results in real order and sequence carry-overs. As noted above, the experimenter will then ordinarily direct his attention to the simple-randomized design formed by the first administration of each treatment condition. Any subsequent measures must be regarded as potentially corrupted, and he must resort to this less precise test. Does this mean he would have been better off in randomizing order and sequence in the first place? No general answer to this question can be phrased. Randomization of these effects adds error variance which would not have been present in the simple-randomized design and reduces the average correlation between the related cells of the Latin square. Conceivably, these combined effects might render the Latin square the less

precise of the two designs, despite the fact the treatment means would be based on N observations in one case, N/a in the other. Order carry-overs tend to run quite large in speech and hearing experiments, often contributing a variance component well over five times the error variance. To randomize such a component and have it contribute to error would result in a less, rather than a more, precise experiment. In view of this genuine possibility, and the fact that the Latin square is rarely used when a sequence carry-over is expected, it appears safe to recommend a counterbalanced handling of treatment order. If it is so controlled or if it is negligible and randomized, the Latin square with two treatment dimensions will hold the same advantage over an orthodox factorial that the order by treatments Latin square has over the simple-randomized design.

Assumptions Underlying the Latin Square

The assumptions which must be met for a valid Latin square analysis are somewhat more difficult to grasp than those of a simple-randomized design. Often even sophisticated experimenters would not care to be pressed on the question of their satisfaction. This in part stems from the fact that the quantities to which the assumptions apply are rarely computed. Hence the typical experimenter can get no feel for this aspect of his data. He cannot make the informal judgments of normality and homogeneity which are often all that are required in the simple-randomized or factorial design. Even if the essential quantities were computed, there is no statistical test in the general case to provide evidence that the assumptions are satisfied. Then, too, the growing literature on the relative unimportance of the assumptions (*1*) has led to a relaxation of the concern which experimenters once felt about small deviations from normality of distribution and homogeneity of variance. Because of these considerations, the typical research report usually avoids all mention of this topic.

If the cells of the Latin square are rearranged so that those of a single group of subjects fall side by side, the design resembles a series of simple 'repeated measures' experiments as indicated below, each by itself a treatments-by-subjects design.

Group I	A_1B_1	A_2B_2	A_3B_3
Group II	A_2B_1	A_3B_2	A_1B_3
Group III	A_3B_1	A_1B_2	A_2B_3

The assumptions in the Latin square are precisely the same as those in this simpler type of design. That is, within each group and from one group to another the treatment-by-subject interaction effects must constitute independent random samples from normally distributed, equally variable populations of such effects.

The assumption of randomness is fully met whenever the entire group of experimental subjects has been chosen at random from the parent population of interest and subdivided randomly to meet the demands of the square. The tenability of this assumption is generally quite clear-cut, and hence needs little discussion here. Deviations from homo-

geneity are considerably harder to ascertain. As noted above, there is no overall test applicable beyond the three-level experiment which will provide statistical evidence on the tenability of the assumption of homogeneity of variance. The components of error which are pooled within any given group cannot be separated so as to meet the requirements of available variance tests. However, it is possible to make a valid partial assessment with respect to those elements which are pooled from the independent groups. This is accomplished by the computation of a separate mean square for treatments-by-subjects interaction for each group. These estimates may then be tested for homogeneity by the Bartlett test or any of the other tests derived for this hypothesis.

Deviation from normality also presents a problem. It could be assessed by available goodness-of-fit tests applied to the computed interaction effects in each cell. However, there is no valid way to combine the results of these tests, since they are not independent, and repeated tests increase the likelihood of at least one significant result. Of course, if all tests proved nonsignificant, this would suggest the assumption is probably sufficiently satisfied. Certainly a large component of the interaction variance is due to errors of observation or measurement, and these are usually found to be approximately normal in distribution. The larger the correlation between cells, the greater the proportion this component will represent. If informal inspection of these distributions indicates no gross asymmetry exists, it seems reasonable to expect that the validity of the *F* test will not be seriously impaired.

These assumptions might be clarified, in the minds of some experimenters, if they were related back to independent group designs. Let it be assumed that the Latin square experiment could have been validly run and analyzed as a simple-randomized or a simple two-dimensional factorial design. That is, let it be assumed that homogeneity of variance and normality of distribution exist throughout all treatment combinations. It then follows that the assumptions of the Latin square analysis also will be satisfied if one additional condition is met: the relationship between any two sets of related measures must be linear and equal throughout the design. (In the diagram noted above, there would be nine product-movement coefficients: cell A_1B_1 against cell A_2B_2, A_1B_1 against A_3B_3, A_2B_2 against A_3B_3, and so on for the other two groups.) So stated, the assumptions are slightly more restrictive than they actually are. It is possible that compensating heterogeneity among the cell variances and among the inter-cell correlations could result in satisfaction of assumptions. However, such offsetting irregularities are probably quite rare.

If the same or equally reliable test forms or measurements are used with each group and the correlation between cells is primarily depressed by test unreliability (errors of measurement), there seems little reason why approximate equality should not generally hold in these correlations. A prime factor leading to inequality, however, is any gross difference in the time interval between measures. Since correlations tend to vary inversely with the time between measures, homogeneity of regression will be more nearly approached when the measurements on each subject

are taken within a short interval of time.

By considering within-cell variances and between-cells correlations, the experimenter may form at least a subjective judgment of the homogeneity of interaction-effect variance. Reasonably similar values should be sufficient evidence upon which to proceed with the analysis. Where the variances are homogeneous but correlations quite variable, a positive bias has been shown for the *F* test (*2*, *3*). That is, there is a tendency for the mean square ratio to exceed the nominal 5 and 1% *F* values too often. In such a situation the experimenter might well compute individual interaction effects and the variance of these effects, cell by cell. The amount of heterogeneity revealed by these estimates might then be referred to the Norton findings (*4*) for at least a rough indication of how the significance level should be modified.

The foregoing discussion has been concerned with the assumptions dictated by the mathematical model underlying the analysis. Probably more important insofar as the interpretation of the experimental results are concerned are the assumptions concerning sequence and order carry-over effects. If these are inappropriately handled and allowed to bias treatment comparisons, a mathematically valid but experimentally invalid analysis can result. Clearly, trustworthy interpretations are as contingent upon the adequate handling of both types of carry-over as upon approximate satisfaction of the mathematical requirements of the *F* tests.

Summary

The purpose is to consider the nature of the Latin square design, the assumptions which underlie it, and its advantages and limitations in speech and hearing research. Among the wide variety of experimental conditions investigated in speech and hearing research, many meet the basic requirements of the Latin square design. Whenever consecutive administration of treatments is experimentally feasible, marked increases in experimental precision can be obtained through this design. The danger of order and sequence carry-overs from one treatment to another is quite real, however. Therefore, the inclusion of order as an experimental variable is generally advisable. Where the effects of consecutive administration accumulate in a simple way common to all treatments, data from the complete square may be utilized to increase precision. Where sequential differences arise, a valid test of treatments effects may still be made using part of the data. The assumptions are closely related to the simple-randomized and factorial analyses, and the effects of deviations from normality and homogeneity may be inferred from research on these analyses.

References

1. Box, G. E. P., Some theorems on quadratic form applied in the study of analysis of variance problems, I. Effects of inequality of variance in the one-way classification. *Ann. math. Statist.*, 1954, 25, 290-302.
2. Gourlay, N., F-test bias for experimental designs in education research. *Psychometrika*, 20, 1955, 227-248.
3. Gourlay, N., F-test bias for experimental designs of the Latin square type. *Psychometrika*, 20, 1955, 273-287.
4. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin, 1953.
5. McNemar, Q., On the use of Latin squares in psychology. *Psychol. Bull.*, 1951, 48, 398-401.

The Conditioned GSR Auditory Speech Threshold

JOSEPH B. CHAIKLIN

The galvanic skin response (GSR) is an autonomic nervous system reaction that consists of a lowering of the electrical resistance of the skin in response to various types of stimulation. Loud sounds, painful or annoying stimuli, motor movements, and internal stimuli are among the events capable of eliciting the GSR. For the last 50 years the GSR has been used in many types of psychological research, but only for the past 10 years has it been employed in a conditioning procedure designed to measure auditory thresholds in subjects who cannot or will not cooperate in standard tests of hearing acuity (*3, 5, 17*). In this procedure the subject is connected into an electric circuit that detects changes in the resistance of his skin to the passage of a weak electric current (Féré effect). The skin resistance changes are amplified and recorded by a graphic recorder, thus providing a permanent record for objective analysis of the subject's responses to stimuli.

During conditioning, the GSR serves as the conditioned response (the response to be conditioned), a pure tone serves as the conditioned stimulus (the stimulus that does not elicit the GSR reliably prior to conditioning), and an unpleasant shock serves as the unconditioned stimulus (the stimulus that elicits the GSR reliably before conditioning). Reinforced pure tones (tone followed by shock) and unreinforced pure tones (tone alone) are presented on a schedule until the unreinforced tones elicit the GSR reliably in accordance with a predetermined conditioning criterion. Once conditioning has been established, intensity generalization enables the tester to assess threshold by presentation of the conditioned stimulus at varying intensity levels. Generalization to tones of different frequencies facilitates threshold assessment throughout the frequency range measured clinically.

Relatively little attention has been directed to the use of speech stimuli in GSR audiometry. Knapp and Gold (*11*) reported using a speech signal as an unconditioned stimulus in a GSR screening technique for the diagnosis of functional hearing loss. Their methodology did not permit systematic threshold measurement since it did not employ conditioning procedures. Ruhm and Carhart (*15*) recently described a conditioned GSR technique for objective speech audiometry. Their methodology involved establishing a conditioned discrimination of a key spondaic word presented in sequence

Joseph B. Chaiklin (Ph.D., Stanford University, 1958) is Assistant Chief, Audiology and Speech Correction Clinic, Veterans Administration Hospital, San Francisco. This article is based on a doctoral dissertation completed under the direction of Professor Hayes A. Newby, and is adapted from a paper presented at the 1958 Convention of the American Speech and Hearing Association, New York. The research was supported by grants from the Speech Correction Fund, the Elks National Foundation, and the Veterans Administration.

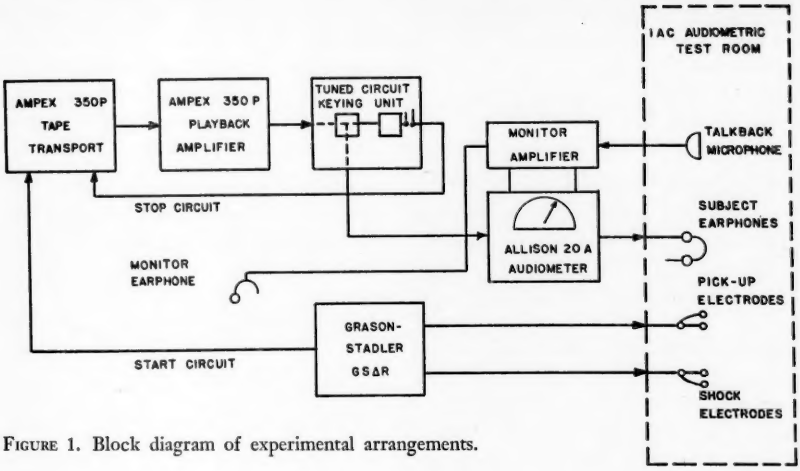

FIGURE 1. Block diagram of experimental arrangements.

among other spondaic words. This was accomplished by reinforcing (with shock) only the key word. The electrodermal speech threshold was defined as the lowest hearing level producing an increase in the magnitude of response to the key-word presentations. The rationale for their technique was that so long as the key word was intelligible to the subject a differential in magnitude of response would be produced. They found that the electrodermal speech threshold was highly correlated with independently measured live-voice speech reception thresholds.

The conditioning procedure employed in the present study was not designed to induce a differential response at a specific threshold level as was the intention in the Ruhm-Carhart research. It was intended, rather, to assess the relative hearing level at which subjects would produce GSRs to a speech stimulus after conditioning was established at a suprathreshold level. This response was designated as the conditioned GSR auditory speech threshold. The speech signal used as the unconditioned stimulus in this experiment was the tape-recorded sentence, 'Now you hear me,' which was spliced into specially prepared tape loops.

Procedure

Purpose. The primary purpose of this study was to investigate the relationships between the conditioned GSR auditory speech threshold and the following voluntary auditory speech thresholds: (a) speech detection threshold (SDT), (b) threshold of perceptibility for running speech (TPRS), and (c) speech reception threshold for CID W-1 spondaic words (SRT W-1).

Subjects. The subjects were 16 males and 14 females who ranged from 23 to 51 years of age with a mean age of 32 years and a median age of 30 years. To be included in the study a subject had to satisfy the following criteria: (a) must

have no known pathology and no greater than a 10-db hearing loss at 500, 1000, or 2000 cps in one ear; (b) must use American English as a first language, and (c) must be able to meet the conditioning criterion for the GSR test.

Equipment. All tests were conducted at the Audiology and Speech Correction Clinic of the Veterans Administration Hospital in San Francisco. The basic test instruments were an Allison Laboratories diagnostic audiometer (Model 20-A) equipped with Telephonics Corporation 10-ohm dynamic earphones (Model TDH39), a Grason-Stadler psychogalvanometer (Model E-664), an Ampex tape recorder (Model 350 P), and a tuned-circuit stimulus keying system. A 10-db nonreactive T-pad was inserted into the line leading to the test earphone to permit threshold assessment in subjects with acute hearing. The audiometer was checked periodically (*1*) with an Allison Audiometer Calibration Unit (Model 300). During test procedures subjects were seated in an Industrial Acoustics Company Audiometric Testing Room (Model 402). The test equipment and the experimenter were located outside the test room. Figure 1 contains a block diagram of the equipment as it was arranged during the experiment.

Stimulus Materials. All speech stimuli were recorded at 7.5 inches per second on the Ampex tape recorder with an Electro-Voice Slimair high-impedance dynamic microphone (Model EV 636) serving as the input source.

The stimulus material for TPRS and SDT was a five-minute prose recording on the subject of ancient history. It was recorded so that most of the material peaked at zero VU during playback. The material was recorded twice in succession so that 10 minutes of continuous running speech was available during speech audiometry.

Lists A, B, and C of the 36 Central Institute for the Deaf (CID) W-1 spondaic words (*8*) served as the stimulus materials for SRT W-1. They were recorded following the running speech section of the stimulus tape. The carrier phrase, 'Say the word,' preceded each spondaic word and a five-second pause occurred between each word and the start of the next carrier phrase.

The sentence, 'Now you hear me,' served as the conditioned stimulus during GSR audiometry. From a population of 20 tape-recorded conditioned stimuli two equated stimuli were selected for use in stimulus tape loops. These loops had the following basic components: (a) a piece of magnetic recording tape containing the conditioned stimulus, (b) a small piece of recording tape containing an inaudible 12000-cps tone, and (c) a piece of leader and timing tape to complete the loop. The 12000-cps keying signal was spliced into the loop preceding the conditioned stimulus section. The tape loops described above were used in an electronic stimulation circuit that utilized the stimulation circuit of the Grason-Stadler psychogalvanometer, the remote control facilities of the Ampex tape recorder, and the tuned circuit keying system which operated the stop circuit of the tape transport. The keying system, tuned to 12000 cycles, rejected all other signals. During the GSR test a tape loop was drawn over the head assembly of the tape recorder each time the stimulation circuit was closed. When the keying signal reached

the playback head the stop circuit of the tape transport was activated. The conditioned stimulus was located immediately after the keying signal, hence the tape loop was in position for the next stimulation sequence. Two equated loops were used to guard against excessive wear on the conditioned stimulus. The plastic take-up and rewind reels of the tape recorder were immobilized by a metal rod hooked over the edge of each reel's outer rim.

Grouping and Order Procedure. The plan of the study called for the administration of five tests (including pure-tone audiometry) of different lengths over a long testing session. To control possible fatigue and order effects inherent in a long testing session, the subjects were divided into two groups (A and B), the tests were arranged in different orders, and a rest period was given. The longest test was the GSR test. In some cases it required an hour. The next longest test (SRT W-1) was about 15 minutes long and the other tests required from 5 to 10 minutes each. The subjects of Group A received the GSR test early in the testing session (after SRT W-1) and the subjects of Group B received the GSR test after all other tests had been completed. Since SRT W-1 served as the reference threshold for GSR conditioning, its position in the test sequence could not be varied beyond set limits. It always preceded the GSR test. For Group A, therefore, SRT W-1 was always the first test and in Group B SRT W-1 was always the last test before the concluding GSR test. The remaining tests–SDT, TPRS, and pure-tone audiometry–were arranged in six permutations. Five subjects were assigned to each permutation in such a fashion that the subjects from both groups were evenly balanced among the permutations.

Threshold Determination. When a subject reported for his appointment he was told that he was going to have a number of hearing tests and that questions about the object of the study would be answered at the conclusion of testing. Written instructions were used and each test was administered in the same way for all subjects. SDT, TPRS, and SRT W-1 were measured in 2-db steps but the limitations of GSR audiometry necessitated sampling in 5-db steps.

SDT. The lowest hearing level at which a subject indicated that he heard speech sounds continuously for five seconds was his SDT. He did not have to understand the speech he heard. The experimenter adjusted the signal intensity in an ascending fashion and the subject used a finger signal to indicate when he heard speech sounds.

TPRS. The determination of TPRS involved finding the lowest hearing level at which a subject reported that he was barely able to understand the running speech presented to him. A descending technique was used for TPRS measurement.

SRT W-1. A subject's SRT W-1 was the lowest hearing level at which he repeated correctly three out of six (that is, 50%) of the tape-recorded spondaic words presented to him. Three estimates were made of each of the voluntary speech thresholds and the median of the three scores was used as the final threshold estimate. A more detailed description of the procedures employed in measuring SDT, TPRS,

and SRT W-1 may be found in another article (*4*).

GSR Audiometry. The GSR test involved the use of a conditioning schedule employing 40 events randomly arranged for hearing level and interval between events. Written instructions were provided to the subject prior to the GSR test. Before conditioning started the unconditioned stimulus was adjusted by discrete-step increments until the subject reported that the shock level was 'distinctly unpleasant' and that he would not want it stronger. After the shock was adjusted the earphones were positioned and the conditioning procedure was started. The reference level during conditioning was SRT W-1 rounded to the nearest 5-db interval. The conditioned stimulus was presented at hearing levels of 5 db, 10 db, and 20 db above the rounded SRT W-1.[1] When the conditioning schedule called for a reinforced presentation of the conditioned stimulus, the interval between the onset of the conditioned stimulus and the onset of the unconditioned stimulus was .5 second (*13, 18*) continuing to the end of the conditioned stimulus (*2*). The conditioned stimulus was approximately one second long. Three intervals were used between events during conditioning. Thirty seconds was the shortest interval, 90 seconds the longest, and 45 seconds an intermediate interval (*10, 16*). Forty per cent partial, random reinforcement was used during conditioning and sampling (*7, 9, 12*). A subject was judged to be conditioned when he produced three successive responses to conditioned stimuli after the fifth event of the schedule had been presented. Waiting until after the fifth event served to protect against premature judgment of conditioning on the basis of unconditioned responses. Latencies (judged from onset of the conditioned stimulus) of one to four seconds and magnitudes one millimeter or greater were used as GSR response criteria in the conditioning process and in analysis of the sampling records. If the conditioning criterion was not met in 30 minutes the subject was dismissed from further testing.

Once the conditioning criterion had been met a five-minute break occurred before the start of the sampling schedule. The sampling schedule used the same reference level and intervals that had been employed in the conditioning schedule. Hearing levels of 5 and 10 db above and below and at the rounded SRT W-1 were sampled. Each of the five hearing levels was sampled five times during the course of the 30-minute sampling procedure. The sampling schedule events were counterbalanced so that the five hearing levels were sampled at points distributed throughout the length of the test. Five control presentations, each consisting of a simulated conditioned stimulus presentation, appeared at specified times in the sampling schedule. An interrupter key opened the circuit to the test earphone during control presentations. This prevented the stimulus from reaching the subject but a marker appeared on the graphic record making it appear as if the subject had received a conditioned stimulus. The control presentations were used as an internal check on the validity of the graphic records. If a subject gave acceptable random responses to two or more of the control

[1] A variable-intensity conditioned stimulus was selected on an arbitrary basis, but a fixed level might have served equally well.

TABLE 1. Means and standard deviations for speech detection threshold (SDT), speech reception threshold (SRT W-1), threshold of perceptibility for running speech (TPRS), and conditioned GSR auditory speech threshold (GSR), N=30.

Measure	*Mean** (*db*)	*Standard Deviation* (*db*)
SDT	16.47	3.45
SRT W-1	25.47	3.85
TPRS	28.93	3.00
GSR	22.83	3.80

*Mean threshold values *re* 0.0002 dyne/cm².

presentations his sampling record was discarded. Only one subject was discarded for this reason.

Analysis of GSR Sampling Records. After the GSR graphic records had been screened by the experimenter to eliminate those that were invalid, they were assigned arbitrary code numbers. The sampling portions of the records were analyzed by three trained judges. The records had on them only numbered events and no information relating to the hearing levels at which stimuli were presented. The judges recorded their analyses of the numbered events on prepared data sheets. The experimenter collated their analyses and was able to determine the threshold hearing level through use of a prepared key. The GSR threshold was the lowest hearing level at which there was majority agreement that the subject responded satisfactorily on three out of five presentations of the conditioned stimulus.

Results

Table 1 summarizes the mean threshold values and standard deviations for each of the thresholds under study. All thresholds are stated in reference to 0.0002 dyne/cm². The mean SDT was 16.47 db; the mean SRT W-1 was 25.47 db, 9 db higher than SDT; the mean TPRS was 28.93 db, 3.46 db higher than SRT W-1; and the mean GSR threshold was 22.83 db.

Table 2 summarizes the results of *t* tests and Pearson product-moment correlations for the comparisons between the GSR threshold and the other thresholds (*6*). An inspection of Table 2 reveals that the GSR threshold was significantly different from each of the other thresholds with values of *t* beyond the .001 level of significance. The GSR threshold was not significantly correlated with TPRS ($r = .21$) and it had low positive correlations with SDT ($r = .47$) and SRT W-1 ($r = .50$), both significant at or beyond the .01 level. It should be noted that the correlation coefficients must be viewed as merely suggestive since the range of performance during GSR audiometry was necessarily limited by the use of 5-db steps and by the normal hearing acuity of the subjects.

It was observed that the GSR threshold had a range of response from approximately 9 db below SRT W-1 to

TABLE 2. Results of the *t* test for related measures and Pearson product-moment correlation coefficients for the intertest comparisons.

Test	*Mean*	*Diff.*	*S.E. Diff.*	*t*	*r*
SDT	16.47	6.36	.69	9.22*	.47†
GSR	22.83				
SRT W-1	25.47	2.64	.71	3.72*	.50**
GSR	22.83				
TPRS	28.93	6.10	.80	7.62*	.21††
GSR	22.83				

*Significant beyond .001 level.
†Significant at .01 level.
**Significant beyond .01 level.
††Not significant at .05 level.

approximately 4 db above SRT W-1. No subject's GSR threshold was higher than 4 db above his SRT W-1. Analysis of the mean test differences between groups A and B revealed values of *t* not significant at the .05 level, indicating that order and fatigue effects did not contribute significantly to test differences. In clinical application of the GSR technique described in this article one would expect the majority of subjects to have GSR thresholds within ±5 db of SRT W-1 with a lesser number responding at lower levels.

Discussion

For most subjects GSR sampling did not go below SDT and for some (seven in all) the lowest hearing level sampled was slightly above SDT. This occurred because GSR sampling went only 10 db below SRT W-1 while SDT was sometimes more than 10 db below SRT W-1. Despite this sampling limitation it is not likely that a significant number of subjects would have responded below SDT. This is borne out by the fact that of the 23 subjects for whom GSR sampling was 2 db above to 9 db below SDT only three had GSR thresholds within 1 db above SDT.

If the GSR threshold criterion were shifted from three out of five acceptable responses to two out of five, then eight subjects rather than three would have had GSR thresholds at the lowest hearing level sampled. One of the five subjects under the two-response-threshold criterion would have had a GSR threshold 4 db below his SDT and the other four would have had GSR thresholds 2 to 4 db above SDT.

It is apparent that the effects of stimulus intensity generalization were strong in the GSR sampling process, especially when one considers the fact that all reinforcement was at levels where the conditioned stimulus was clearly intelligible. Yet some subjects produced acceptable GSRs at levels where the conditioned stimulus was not intelligible to them. These generalization effects would have to be recognized in any clinical application of the GSR technique described in this article.

An analysis of the mean number of responses at each sampling level suggests that some form of intensity dynamism operated in response determination. At 10 db below SRT W-1 the mean number of responses was .90. At 5 db below SRT W-1 the mean number of responses rose to 2.67. At SRT W-1 the mean number of responses was 3.67 and at 5 and 10 db above SRT W-1 the mean number of responses was 4.77 and 4.60, respectively. The increase in response frequency appears to be negatively accelerated, with maximum response occurring once SRT W-1 is passed.

An incidental finding was that females appeared more difficult to condition than males. Of the 28 female subjects who took the GSR test 50% (14 subjects) met the conditioning criterion while 89% (16 subjects) of 18 males tested met the conditioning criterion. A similar sex differential was recently reported by Rosenblüt, Bilger, and Goldstein (*14*).

Summary

Sixteen male and 14 female subjects with normal hearing were conditioned to produce GSRs to a conditioned stimulus consisting of the tape-recorded sentence 'Now you hear me.' After conditioning was established, threshold was assessed by systematic sampling with

the conditioned stimulus. The lowest hearing level at which three GSRs were elicited in five presentations was designated the conditioned GSR auditory speech threshold (GSR threshold). Speech detection threshold (SDT), speech reception threshold for CID W-1 spondaic words (SRT W-1), and threshold of perceptibility for running speech (TPRS) were measured as part of the experimental procedure.

Statistical comparisons between the GSR threshold and each of the other thresholds revealed significant differences (values of t beyond the .001 level) and low positive correlations (GSR-TPRS = .21, GSR-SDT = .47, GSR-SRT W-1 = .50) for all comparisons. Even though all reinforcement was at levels above threshold of intelligibility, the effects of intensity generalization were sufficiently strong to cause some subjects to produce GSRs at levels where the conditioned stimulus was not intelligible to them. The majority of subjects had GSR thresholds within ±5 db of SRT W-1. The effects of generalization, however, would have to be recognized in any clinical application of the GSR procedures employed in this study. An incidental finding was that women appeared more difficult to condition than men.

References

1. American Standards Association, *American standard specification for speech audiometers.* Report Z24.13-1953, Section 3.12.
2. Bitterman, M. E., Reed, P., and Krauskopf, J., The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.*, 65, 1952, 256-262.
3. Bordley, J. E., and Hardy, W. G., A study in objective audiometry with the use of a psychogalvanometric response. *Ann. Otol. Rhinol. Laryngol.*, 58, 1949, 751-760.
4. Chaiklin, J. B., The relation among three selected auditory speech thresholds. *J. Speech Hearing Res.*, 2, 1959, 237-243.
5. Doerfler, L. G., and McClure, Catherine T., The measurement of hearing loss in adults by galvanic skin response. *J. Speech Hearing Dis.*, 19, 1954, 184-189.
6. Edwards, A. L., *Statistical Analysis.* New York: Rinehart, 1946.
7. Grant, D. A., Meyer, D. R., and Hake, H. W., Proportional reinforcement and extinction of the conditioned GSR. *J. gen. Psychol.*, 42, 1950, 97-101.
8. Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, Elizabeth G., Eldert, Elizabeth, and Benson, R. W., Development of materials for speech audiometry. *J. Speech Hearing Dis.*, 17, 1952, 321-337.
9. Humphreys, L. G., Extinction of conditioned psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 27, 1940, 71-75.
10. Kimble, G. A., Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 37, 1947, 1-15.
11. Knapp, P. H., and Gold, B. H., The galvanic skin response and diagnosis of hearing disorders. *Psychosom. Med.*, 12, 1950, 6-22.
12. Meritser, C. L., and Doerfler, L. G., The conditioned galvanic skin response under two modes of reinforcement. *J. Speech Hearing Dis.*, 19, 1954, 350-359.
13. Moeller, G., The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 48, 1954, 162-166.
14. Rosenblüt, B., Bilger, R. C., and Goldstein, R., Electrophysiologic responses to sound as a function of intensity, EEG pattern and sex. *J. Speech Hearing Res.*, 2, 1959, 28-39.
15. Ruhm, H. B., and Carhart, R., Objective speech audiometry: A new method based on electrodermal response. *J. Speech Hearing Res.*, 1, 1958, 169-178.
16. Spence, K. W., and Norris, Eugenia B., Eyelid conditioning as a function of the inter-trial interval. *J. exp. Psychol.*, 40, 1950, 716-720.
17. Stewart, K. C., Some basic considerations in applying the GSR technique to the measurement of auditory sensitivity. *J. Speech Hearing Dis.*, 19, 1954, 174-183.
18. White, C. T., and Schlosberg, H., Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 43, 1952, 357-362.

The Relation Among Three Selected Auditory Speech Thresholds

JOSEPH B. CHAIKLIN

Speech audiometry embraces the clinical procedures used in measuring auditory response to speech stimuli. Speech reception threshold (or threshold of intelligibility) and speech sound discrimination are the stimulus dimensions assessed most frequently in clinical practice. Speech reception threshold (SRT) is usually defined as the lowest hearing level at which a person repeats correctly 50% of the spondaic words heard by him. Speech sound discrimination (intelligibility function) is usually measured with lists of monosyllabic words phonetically balanced to include speech sounds in their approximate frequency of occurrence in daily conversation. The words are presented 30 to 40 db above SRT and the person under test says or writes the words as he hears them. The history of the development of SRT and intelligibility tests is well-documented elsewhere (*2, 6, 8, 9, 10*). The spondaic words of CID (Central Institute for the Deaf) Auditory Test W-1 and the phonetically balanced words of CID Auditory Test W-22 are the test materials in widest use at this time (*9*).

In some instances speech audiometry includes measures that employ running speech (or connected discourse) as a stimulus. The thresholds of detectability, perceptibility, and intelligibility described by Egan (*6*) employ running speech as a stimulus. Thresholds of comfort and discomfort for speech are frequently included in speech audiometry as indexes of recruitment (*11*).

The published reports relating to speech thresholds provide rather incomplete data, particularly for the thresholds of detectability and perceptibility described by Egan. Thurlow and others (*13*) found threshold of detectability was 9 db lower than threshold for Psycho-Acoustics Laboratory (PAL) Auditory Test Number 9 (spondaic words) when they used 14 normal ears, but only 5.05 db lower for 237 hard-of-hearing ears. Hirsh (*8*) reported that the threshold of detectability for speech has a mean value of about 15 db and SRT for spondaic words (unspecified) is approximately 25 db (both *re* 0.0002 dyne/cm^2). Preliminary work at the Central Institute for the Deaf (*9*) indicated an approximate threshold of 21 db (*re* 0.0002 dyne/cm^2) for CID Auditory Test W-1 with inexperienced

Joseph B. Chaiklin (Ph.D., Stanford University, 1958) is Assistant Chief, Audiology and Speech Correction Clinic, Veterans Administration Hospital, San Francisco. This article is based on a portion of a doctoral dissertation completed under the direction of Professor Hayes A. Newby. The research was supported by grants from the Speech Correction Fund, the Elks National Foundation, and the Veterans Administration.

normal-hearing listeners. Corso (*4*) later established a threshold of 18.55 db (*re* 0.0002 dyne/cm²) for CID Auditory Test W-2 using 278 normal ears.

Falconer and Davis (7) studied the threshold of intelligibility for connected discourse (TICD) as a possible clinical tool for quick assessment of threshold for speech stimuli. They used a recording of Fulton Lewis, Jr., as a stimulus passage and for 50 normal ears found a mean TICD of 23.23 db (*re* 0.0002 dyne/cm²), in comparison with a mean threshold of 22.46 db (*re* 0.0002 dyne/cm²) for Auditory Test Number 9. They concluded that TICD, as administered in their study, is a reliable and valid threshold test.

The purpose of the present study was to investigate on an exploratory basis the relationships among speech reception threshold for CID W-1 spondaic words (SRT W-1), speech detection threshold (SDT), and threshold of perceptibility for running speech (TPRS).

Procedure

Subjects. The subjects were 16 males and 14 females who ranged from 23 to 51 years of age with a mean age of 30 years, spoke American English as a first language, and had at least one ear with pure-tone thresholds at 500, 1000, and 2000 cps no higher than the 10-db hearing level.

Equipment. The basic test instrument was an Allison Laboratories diagnostic audiometer (Model 20A) equipped with Telephonics Corporation 10-ohm dynamic earphones (Model TDH 39). Speech stimuli were recorded and played back on an Ampex Corporation tape recorder (Model 350 P). The equipment has been described in greater detail in another article (*3*).

Calibration Checks. Calibration checks of the Allison audiometer speech circuit were performed periodically (*1*) with an Allison Laboratories Audiometer calibration unit (Model 300). The sound pressure output at the earphone was 24 db (*re* 0.0002 dyne/cm²) when a 1000-cps tone was set at zero VU on the Allison audiometer VU meter and the hearing loss dial was set at 0 db. During all threshold measurements an additional 10 db of attenuation was inserted into the test earphone line by means of a nonreactive T-pad. This was necessary to provide an adequate range of attenuation below normal threshold for spondaic words.

Stimulus Materials. The same speaker (the experimenter) recorded all stimulus materials. The stimulus material for SDT and TPRS was a tape-recorded prose selection concerning ancient history. The running speech (connected discourse) script, which took five minutes to read, was read twice in succession providing 10 minutes of uninterrupted stimulus material. The script was read in such a fashion that most of the material peaked at zero VU. Lists A, B, and C of the 36 CID W-1 spondaic words (*9*) served as the stimulus materials for SRT W-1. They were recorded following the recording of the running speech section of the stimulus tape. The carrier phrase, 'Say the word,' peaked at zero VU, preceded each spondaic word. No special effort was made to peak the spondaic words at zero VU. A five-second interval elapsed between each spondaic word and the start of the succeeding carrier

phrase. The sustained vowel [ɑ] was recorded at zero VU at the start of the stimulus tape for use in adjusting input level at the start of each testing session. A single setting brought the carrier phrases and most of running speech to zero VU.

Threshold Assessment. The threshold tests used in this study were administered in the same fashion for all subjects. To control possible systematic order and fatigue effects a counterbalanced design was used. The counterbalancing procedure has been described previously (*3*). Three trials were made for each of the thresholds under study and the median of the trials was used as the final threshold estimate. Written instructions were employed which the experimenter read to the subject after the subject had read them to himself. The right ear was tested unless the subject reported, during routine questioning, that his left ear was his better ear.

The following instructions were used for SDT:

> Each time you hear something which sounds like speech, signal by raising your right index finger. Keep your finger raised as long as you hear something which sounds like speech. You do not have to understand the speech which you hear, but you should recognize the sounds as being speech—that is, the sound of someone talking. We will make several trials for this measurement. Be certain to put your finger down when you are not hearing something which sounds like speech.

At the start of each SDT trial the running speech stimulus was fully attenuated. The signal was increased in 2-db steps with five seconds of stimulation at each step. If the subject did not respond after five seconds the next higher step was tried. This procedure continued up to the level where the subject signalled properly for threshold determination. For each trial SDT was judged to be the lowest hearing level at which a subject indicated that he heard speech sounds continuously for five seconds. A five-second criterion for threshold estimate was used in an attempt to reduce the number of unreliable threshold estimates resulting from momentary reactions to stimulus intensity peaks not representative of the average intensity level of the speech stimulus.

The following instructions were used for TPRS:

> You are going to hear a recording of a man talking. The loudness of his voice will be decreased gradually. When his voice reaches a point where you can barely understand what he is talking about, *raise* your right index finger. It is a point where you may miss some of the words he is using, but where you still have a fair idea of what is being said. Keep your finger raised *as long as you can barely understand what is being said.* If I go below the point where you can barely understand what is being said, signal by lowering your right index finger. We will make several trials for this measurement. At the start of each trial I will make the speech loud enough to be understood easily. The only time your finger should be up is during the period when you are barely able to understand the speaker.

The test was started with the running speech signal set at 35 db above the normal reference level (24 db *re* 0.0002 dyne/cm^2). The signal was attenuated at a slow rate (about 2 db per second) until the subject signalled that he could barely understand the recorded material. If, for five seconds, he continued to indicate that he barely understood the recording, the next lower 2-db step was tried. The same five-second criterion was applied for each succeeding 2-db decrease in signal strength. The signal

was attenuated in this fashion until the subject lowered his finger to indicate that he no longer understood the stimulus passage. The lowest 2-db hearing level at which the subject barely understood the recording for five seconds represented TPRS.

The measurement of SRT W-1 was preceded by the following instructions:

> You are going to hear a series of two-syllable words during the next few minutes. Your task will be to repeat the words as you hear them. At times the words will get very faint. When they get faint please try to guess at them. There will be no words longer than two syllables, and there will be no words shorter than two syllables. I am going to let you look at all the words which you will hear so that you can become familiar with them. There will be several trials for this measurement. At the start of each trial the words will be loud enough to be heard easily. After you have looked at the words, return them to me and the test will begin.

After the instructions were read, the subject was handed a stack of 36 cards. On each card there was typed one of the 36 CID W-1 spondaic words. After the subject had looked at each of the words the test was begun. At the start of each trial the spondaic words were set at 35 db above the normal reference level (24 db *re* 0.0002 dyne/cm^2). The signal was attenuated in 5-db steps with one word being presented at each step until the subject failed to repeat a word correctly. The signal was then increased 5 or 6 db to an even step on the hearing-loss dial. At each succeeding 2-db decrement spondaic words were presented until (a) the subject repeated three words correctly without exceeding a maximum of six presentations, or (b) the subject missed four words without exceeding a maximum of six presentations. At each level, therefore, there were presented a minimum of three words (that is, three correct) or a maximum of six words (that is, three correct and three incorrect, or two correct and four incorrect). The threshold standard implicit in this method was that of 50% correct (three out of six). Thus at each level when the minimum limits described above were met, the experimenter moved to the next lower 2-db step. A tally of correct and incorrect responses was made as the test progressed. When four words were missed at each of three consecutive levels sampling was discontinued. This criterion for sampling discontinuance was used as a protection against premature acceptance of an early 50%-level as the lowest hearing level at which 50% of the spondaic words could be repeated correctly. For each of the three trials SRT W-1 was defined as the lowest hearing level at which the subject repeated three words correctly. It generally required about 15 minutes to complete the three trials for SRT W-1.

Results

Table 1 summarizes the mean threshold values and standard deviations for each test. The mean SDT was 16.47

TABLE 1. Means and standard deviations for speech detection threshold (SDT), speech reception threshold (SRT W-1), and threshold of perceptibility for running speech (TPRS), N = 30.

Measure	*Mean** (*db*)	*Standard Deviation* (*db*)
SDT	16.47	3.45
SRT W-1	25.47	3.85
TPRS	28.93	3.00

*Mean threshold values *re* 0.0002 dyne/cm^2.

TABLE 2. Results of the *t* test for related measures and Pearson product-moment correlation coefficients for the intertest comparisons.

Test	*Mean (db)*	*Diff. (db)*	*S.E. Diff.*	*t*	*r*
SRT W-1	16.47	3.46	.73	4.74*	.27†
TPRS	28.93				
SRT W-1	25.47	9.00	.66	13.64*	.51**
SDT	16.47				
SDT	16.47	12.46	.65	19.17*	.35†
TPRS	28.93				

*Significant beyond .001 level.
†Not significant at .05 level.
**Significant beyond .01 level.

(*re* 0.0002 dyne/cm²), the mean SRT W-1 was 25.47 db, and the mean TPRS was 28.93 db. Pearson product-moment correlations and *t* tests for related measures (*4*) were employed in statistical analyses which are summarized in Table 2. Values of *t* significant beyond the .001 level were found for each of the mean comparisons, indicating significant mean threshold differences among the tests. The correlational analyses produced low positive values of *r* for all three comparisons. The .51 correlation between SRT W-1 and SDT was significant (P = .01), but the SRT W-1-TPRS correlation ($r = .27$) and the TPRS-SDT correlation ($r = .35$) were not significant. The low correlations between thresholds raised the question of whether each of the three threshold tests might be unreliable enough to contribute to the low correlations between tests. To check on this possibility, average intercorrelations (*12*) were computed among the three trials that were made for each test. The average intercorrelations among trials yielded an *r* of .82 for trials of TPRS, an *r* of .87 for trials of SDT, and an *r* of .77 for trials of SRT W-1. These findings indicate that the separate tests were relatively reliable measures, hence the low correlations between tests may be presumed to have been primarily a function of differences between the tests rather than a function of their unreliability.

Discussion

The results summarized above suggest that in the average case the instructions for each test served to produce significant differences among the thresholds. The relatively low correlations observed between tests are consistent with the variability in test relationships observed from subject to subject. For example, while the mean SDT was 9 db lower than the mean SRT W-1, several subjects had SDT and SRT W-1 determinations that were within 2 db of each other. Yet there were other instances in which these thresholds were 12 to 16 db apart. Sometimes SRT W-1 was within 2 db of TPRS but in several cases TPRS was 8 to 10 db higher than SRT W-1. Both SDT and TPRS evidenced little tendency to approximate each other. It was observed that TPRS was always higher than SDT, yet even here individual variation was apparent with differences between the two thresholds ranging from 6 to 20 db.

Comparison with Previous Findings. Egan (*6*) reported TPRS was 8 db above SDT as contrasted with the 12.47-db gap between SDT and TPRS found in the present study. Hirsh's (*8*) reported mean SDT of 15 db (*re* 0.0002 dyne/cm²) agrees closely, however,

with the mean SDT of 16.47 db reported above. The 9-db difference found between SDT and SRT W-1 agrees identically with the difference between SDT and threshold for spondaic words (Auditory Test No. 9) found by Thurlow and others (*13*) for normal ears. The mean SRT W-1 of 25.47 db reported in the present study agrees well with Hirsh's (*8*) general estimate of threshold for spondaic words (25 db *re* 0.0002 dyne/cm²) but it is higher than the tentative threshold of 21 db (*re* 0.0002 dyne/cm²) stated by Hirsh and others (*9*) for the CID recordings of CID Auditory Test W-1, and higher than the 18.55 db (*re* 0.0002 dyne/cm²) threshold Corso (*4*) found for CID Auditory Test W-2. Falconer and Davis' (7) finding that TICD agrees closely with the threshold for PAL Auditory Test No. 9 (spondaic words) is not consistent with the findings of the present study. Since TICD is the level at which a subject believes he understands all of the running speech material without perceptible effort, it is by definition a higher threshold than TPRS (where the subject is barely able to understand the stimulus passage). In the present study the mean TPRS was 3.46 db higher than SRT W-1 and while TICD was not measured, Egan's (*6*) estimate that it is about 4 db above TPRS appears reasonably accurate. One can estimate, thus, that if TICD had been measured in the present study it would have been approximately 7 db above SRT W-1.

Differences in stimulus materials, sample composition, or methodology may account for some of the differences noted above. For example, Corso tested 139 subjects, 18-24 years old, with W-2 recordings in an anechoic chamber, while in the present study 30 subjects, 18-51 years old, were tested with different recordings, in a room not having anechoic properties. It appears that there is fairly good agreement that the normal threshold for spondaic words lies somewhere between 18-25 db (*re* 0.0002 dyne/cm²), which is consistent with the 22 db (*re* 0.0002 dyne/cm²) normal threshold specified for speech audiometers by the American Standards Association (*1*). The threshold relationships among SRT, TICD, and TPRS may be clarified by further research with larger populations.

Summary

Sixteen male and 14 female subjects with normal hearing received the following threshold tests: (a) speech detection threshold (SDT), (b) speech reception threshold for CID W-1 spondaic words (SRT W-1), and (c) threshold of perceptibility for running speech. The stimulus materials, which were tape recorded by one speaker (the experimenter), consisted of a sample of running speech for SDT and TPRS measurements and lists A, B, and C of the CID Auditory W-1 spondaic words for SRT W-1.

The mean SDT was 16.47 db, the mean SRT W-1 was 25.47 db, and the mean TPRS was 28.93 db (all *re* 0.0002 dyne/cm²). Statistical analyses revealed significant differences (values of *t* beyond the .001 level) and low positive correlations for all threshold comparisons. A discussion of threshold data from previous reports points out several areas of difference which might be resolved by further research.

References

1. American Standards Association, *American standard specification for speech audiometers.* Report Z24.13-1953, Section 3.12.
2. Black, J. W., The origin and nature of the studies (in Studies in speech intelligibility, a program of war-time research). *Speech Monogr.*, 13, 2, 1946, 1-3.
3. Chaiklin, J. B., The conditioned GSR auditory speech threshold. *J. Speech Hearing Res.*, 2, 1959, 229-236.
4. Corso, J. F., Confirmation of the normal threshold for speech on C. I. D. Auditory Test W-2. *J. acoust. Soc. Amer.* 29, 1957, 368-370.
5. Edwards, A. L., *Statistical Analysis.* New York: Rinehart, 1946.
6. Egan, J. P., Articulation testing methods. *Laryngoscope,* 58, 1948, 955-991.
7. Falconer, G. A., and Davis, H., The intelligibility of connected discourse as a test for the "threshold for speech." *Laryngoscope,* 57, 1947, 581-595.
8. Hirsh, I. J., *The Measurement of Hearing.* New York: McGraw-Hill, 1952.
9. Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, Elizabeth G., Eldert, Elizabeth, and Benson, R. W., Development of materials for speech audiometry. *J. Speech Hearing Dis.*, 17, 1952, 321-337.
10. Hudgins, C. V., Hawkins, J. E., Karlin, J. E., and Stevens, S. S., The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope,* 57, 1947, 57-89.
11. Newby, H. A., *Audiology; Principles and Practice.* New York: Appleton-Century-Crofts, 1958.
12. Peters, C. C., and Van Voorhis, W. R., *Statistical Procedures and their Mathematical Bases.* New York: McGraw-Hill, 1940.
13. Thurlow, W. R., Silverman, S. R., Davis, H., and Walsh, T. E., A Statistical study of auditory tests in relation to the fenestration operation. *Laryngoscope,* 58, 1948, 43-66.

A Phonetic Study of Misarticulation of /r/

JAMES F. CURTIS

JAMES C. HARDY

Considerable interest has been shown in the phenomenon of inconsistent misarticulation in the speech of children, particularly among children who exhibit deviant speech without apparent organic causation. Several authors have suggested that occasional correct productions of a speech sound which is usually misarticulated may be used to advantage in helping a child learn to articulate that sound in the correct manner consistently. Several studies, for example Amidon (*1*), Wilson (*15*), and Roe and Milisen (*11*), have shown that correct utterance of sounds which are frequently misarticulated is common in the speech patterns of young children who exhibit defective articulation. However, only when intensive studies of specific phonemes were made, such as the studies of /s/ by Nelson (*10*) and Hale (*3*), and the study of /r/ by Buck (*2*), did evidence begin to accumulate showing that there is a systematic association between particular phonetic contexts and greater frequency of correct productions of these sounds. Spriestersbach and Curtis (*12*) summarize data from several studies which point to this general conclusion. If, indeed, this seemingly inconsistent articulatory behavior is actually systematic and lawful, it is desirable to determine not only the presence of such phenomena in the speech of children having articulatory defects, but also the factors which bring about the systematic occurrence of correct productions.

Of particular interest in relation to the present study is Buck's investigation of /r/. As indicated, Buck's data suggest that children articulate /r/ sounds differently in different phonetic contexts. In addition, his data show rather different articulatory response patterns for the vowel /r/ sounds than for the consonant glide /r/. However, in the view of the present writers, Buck's procedure placed definite limitations on the extent to which these relationships could be explored. It seemed possible that with certain modifications in procedure from those employed by Buck a more complete description of the /r/ responses of children judged to have /r/ defects might be obtained. Accordingly, it was the principal purpose of this study to extend the data concern-

James F. Curtis (Ph.D., University of Iowa, 1942) is Professor and Head, Department of Speech Pathology and Audiology, University of Iowa. James C. Hardy (M.A., University of Iowa, 1957) is Supervisor of Speech Therapy, Hospital School for Severely Handicapped Children, University of Iowa. A portion of the results of this study was presented in a paper at the 1958 Convention of the American Speech and Hearing Association, New York.

ing /r/ by employing what were believed to be improved procedures of observation and classification.

Procedure

Modification of Procedures Previously Used. The procedures employed by Buck were modified in the following ways.

a. Buck classified the position in which each sound occurred as 'initial,' 'medial,' and 'final' in relation to the word in which it occurred without consideration to the relation of the sound to the syllable of which it was a part. The literature on articulatory disorders and articulation testing shows this to be a common practice. However, there would appear to be good reason for the relationship of a sound to its syllable to be considered more important than its position in a word. For example, Stetson's (*13*) analysis of articulatory dynamics, supported by considerable experimental data, appears to indicate that a consonant's function as either an initiating or a terminating element of a syllable may be more significant than its position in a word. Classifying a sound as 'medial' in a word ignores this possible difference in function, since a consonant in the 'medial' position may be either initiating or terminating with respect to a syllable of that word.[1] Therefore, the present study did not divide sounds according to position within words, but rather in relation to the syllabic element which is, of course, usually a vowel. Hence, the classification here used with respect to consonant /r/s was based on whether the sound occurred as a prevocalic, or a postvocalic sound within the syllable.

b. A second difference with respect to the analysis and classification of sounds between Buck's investigation and the present study refers to the classification of /r/ in such words as *car* and *board*. Phoneticians differ in the symbols employed for these /r/s. Thomas (*14*) and Kenyon and Knott (*8*) use the same symbol for this case as for the initiating glide sound in *read*. This implies that both function as consonants. Hultzen (*5*), on the other hand, has argued that these /r/ sounds are more accurately considered as vowel elements of diphthongs, and Buck's data contained indications that the responses of his subjects to these /r/s were more like their responses to the vowel /r/s of *bird* and *butter* than they were like their glide consonant (prevocalic) /r/ responses. Hence, in this study these postvocalic and preconsonantal /r/s were transcribed as vowels and classified as elements of diphthongs.

c. A third departure of significance concerned the manner in which articulatory errors were described. Buck classified errors according to the usual division of omissions, substitutions, and distortions. However, classifying a misarticulation as a distortion reveals little concerning the specific response made. It seemed likely that a more complete description which would indicate what sounds were actually produced, or the nature of the phonetic variation that

[1]The 'medial' position classification led Buck into particular difficulty with respect to certain sound sequences that were termed /r/ blends. For example, in such a word as *bedroom*, the [dr] sequence was considered as a blend containing /r/. More careful analysis will indicate that these two sound elements probably belong to different syllables and do not constitute a blend in the same sense as does the [tr] sequence in a word like *track*.

occurred when the sounds were distorted, would give additional insight into the articulatory behavior of the subjects. Hence, instead of employing the usual omission, substitution, and distortion classifications, the present study attempted, by means of a system of close phonetic transcription, to symbolize the misarticulations that the subjects produced. Thus, errors were classified either as omissions or substitutions for which the error sound was transcribed.

d. Lastly, to obtain more information concerning intrasubject trends of systematic misarticulation, the speech sample elicited from the subjects in this study included a greater number of /r/ attempts than did the sample obtained by Buck.

Subjects. The subjects for this study were 30 children between the ages of 5 years, 6 months, and 8 years, 6 months, who had been diagnosed by a speech correctionist as having functional misarticulation, /r/ being one of the problem phonemes in each case. All of the subjects could be considered within the normal range of intelligence as evidenced by satisfactory school progress. All of the subjects were from sufficiently similar speech environments to warrant one dialectal standard being used in judging their /r/ errors.

Testing Procedure. The test materials consisted of a series of picture cards designed to elicit 175 words containing a total of 195 examples of /r/. The subjects were presented with the 175 test cards in an orthodox manner of articulation testing making no use of oral stimultation. If, on first presentation of a card, a child did not say the desired word even with the help of a starter question, the word was passed over. After the complete set of 175 cards had been presented, those pictures which had failed to elicit the expected words were again shown to the child. If the desired response was still not forthcoming, no further effort was made to elicit that particular word. For such items a 'no response' notation was entered on the data sheet.

Recording and Preparation of Tapes for Transcription. The testing sessions were recorded on plastic base magnetic recording tape by means of a Magnacorder PT6-A4 recording unit with an associated PT6-J amplifier and Electro-Voice Model 731 microphone. Tape speed was 7.5 inches per second. By means of a re-recording process involving the use of two tape recorders, an Ampex Model 403-2 and the Magnacorder equipment mentioned previously, the responses of subjects were transferred to another tape with the extraneous speech of the examiner and subjects deleted. During this process, adjustments were made to equalize the recorded level of the words on the re-recorded tape.

To avoid the possibility that continuous listening to a single subject's responses might lead the observers to expect certain patterns of misarticulation, the following procedure was used to intermix and scramble the responses of different subjects. The re-recorded tapes were cut into segments containing from one to four words. After the word and subject number had been written on the back of each piece of tape with India ink, the segments were mixed and spliced together in random order. The resulting transcription tapes carried about 5000 isolated words.

Method of Describing the Speech Errors. Previous mention has been made of the fact that a system of narrow phonetic transcription was devised to describe the articulations composing the test words as spoken by the subjects. In order for the reader to more easily interpret the discussion and data presented below, explanation of this system of transcription is presented.

Most systems of narrow phonetic transcription make use of diacritics having a physiological reference. For example, notations are used which indicate that the tongue is lowered, backed, raised, or fronted. This use of diacritics is vulnerable to the objection that it attempts to infer details of physiological movement that are not directly observed. The accuracy with which such inferences may be made is subject to question. What one can observe are acoustical events. Accordingly, the system adopted for this study attempted to describe variations in the acoustically perceived utterances without reference to the physiological variations which may have accompanied them.

The phonemic symbols of the International Phonetic Association as given by Kenyon and Knott (*8*) were used as the basic symbols of the system. Modifying symbols were employed in relation to the families of sounds represented by the basic symbols in accordance with the actual event which was heard. For instance, if the [t] of the word [prɛzənt] was perceived to be aspirated, the diacritic [h] was placed to the right of and slightly above the symbol /t/, thus forming the allophone [t^h]. If, in the case of vowels, the /ɪ/ in [tɝkɪ] was perceived to contain elements of both the /ɪ/ and /i/, the /i/ symbol was placed just above the /ɪ/ symbol, thus indicating the allophone, or event, $\left[\begin{smallmatrix} i \\ ɪ \end{smallmatrix}\right]$. If the observer judged the sound to contain a larger proportion of the elements of /ɪ/ than of /i/, the /ɪ/ would appear larger and be placed under the symbol of the minor element. Thus, the transcription system provided for gradation between any two elements in the phonemic repertoire of the observers.

One decision was made which involved a major deviation from the symbols as presented by Kenyon and Knott. Those authors use the symbol [r] to represent the consonant glide /r/, [ɝ] to represent the stressed vocalic /r/, and [ɚ] to represent the unstressed vocalic /r/. However, as mentioned earlier they use the consonant symbol [r] to transcribe the postvocalic /r/ in such words as *card* and *bear*. For reasons previously discussed, in this study these postvocalic and preconsonantal /r/s were considered to be vowel sounds, rather than consonants. Hence, the decision was made to transcribe the postvocalic /r/s with the symbol [ɚ].

Observers and Transcription Procedure. Two observers, both of whom had extensive experience in testing the articulation skills of children, transcribed the words on the transcription tapes. Both observers had worked as public school speech therapists, were graduate students in speech pathology and audiology, and had participated in a graduate seminar in advanced phonetics. In order to eliminate as many differences as possible in the two observers' transcriptions, several practice

TABLE 1. Description and examples of the three types of /r/ employed in organizing the data.

Types of /r/	*Description of Types of /r/*	*Examples*
/r/	The consonant glide /r/ which occurs in the prevocalic position as a single and in blends	*r*abbit *tr*ain
/ɝ/	The stressed vocalic /r/	bi*rd* chu*rch*
/ɚ/	The unstressed vocalic /r/ as a syllabic element and as a nonsyllabic element of a diphthong	mustar*d* wate*r* bea*r* boar*d*

sessions were held during which the observers transcribed recordings of articulatory defective speech. This practice was continued until it was judged that the two observers were employing the transcription system with sufficient consistency and agreement to justify starting the transcription of the experimental tapes.

The transcription tapes were played from the Magnacorder to which was connected a model 210A Scott amplifier and a Jensen 15″ coaxial speaker. One observer read the number appearing on the tape as it approached the playback head and then stopped the recorder. Both observers wrote this number on a ruled pad. Next the recorder was turned on until the word was played when it was again stopped. The observers then independently transcribed the word. A word was replayed until each observer was satisfied with his transcription.

Results

Organization of Data. Classification of the data involved two major steps: (a) the various examples of the /r/ phoneme included in the speech tests had to be classified according to types of /r/, and (b) the various phonetic contexts had to be organized in such

TABLE 2. Description and examples of the phonetic contexts in which each type of /r/ occurred.

Symbol for Phonetic Context	*Description of Phonetic Context*	*Examples*
/rV/	The glide consonant /r/ which occurs as a 'single' in the prevocalic position, followed by a vowel, /V/	*r*abbit *r*ock
/Cr/	The glide consonant /r/ which occurs in 'blends' in the prevocalic position, preceded by a consonant, /C/	pre*s*ent *tr*ain
/ɝ/	The stressed vocalic /r/	bi*rd* sh*irt*
/Vɚ/	The unstressed vocalic /r/ as it occurs after a vowel, /V/	fi*re* bea*r*
/VɚC/	The unstressed vocalic /r/ as it occurs in a postvocalic, preconsonant context, with /V/ standing for a vowel and /C/ a consonant	ho*rn* fo*rk*
/Cɚ/	The unstressed vocalic /r/ as it occurs as a postconsonant sound, with /C/ representing the consonant	wa*t*e*r* ri*v*e*r*
/CɚC/	The unstressed vocalic /r/ as it occurs between two consonants	mus*t*a*rd* sci*ss*o*rs*

a manner as to reveal possible systematic influences of adjacent sounds in facilitating or inhibiting correct production of /r/.

Types of /r/. Three classifications differentiating between types of /r/ were employed in this study. They were the consonant glide /r/, the stressed vocalic /r/, and the unstressed vocalic /r/. Description and examples of these classifications appear in Table 1.

Organization by Phonetic Context. When the three types of /r/ are considered in terms of phonetic contexts in which they occur, each type falls into certain phonetic environments described in Table 2.

As previously explained, this classification does not make the usual division of sounds into initial, medial, and final positions, but takes account, instead, of the function of sounds in syllables. One difficulty with the present classification arises from the fact that instances occur for which there is real uncertainty concerning the appropriate division of a word into syllables. For example, in the word *apron*, it is possible for the word to be uttered so that the combination [pr] is produced as a consonant blend which initiates the second syllable. On the other hand, the two consonant elements [p] and [r] may be uttered so that they would be more appropriately considered as belonging to different syllables. Twenty-nine of the 195 examples of /r/ of this study occurred in phonetic contexts of this type. That is, the phonetic context of the /r/ in relation to a syllable could not be unequivocally determined on *a priori* grounds. These 29 examples of /r/ were included in a separate category of test responses, which were termed intersyllabic /r/ contexts, and were considered only for comparative purposes with the main body of the data. These intersyllabic contexts will be discussed in a later section.

The remaining observations which constitute the main body of data concern those occurrences of /r/ whose position in regard to the syllable in which they occurred could be determined because they were (a) a consonant glide /r/ and initiated the test word, (b) a vocalic element constituting either the syllabic element of the syllable, or the second element of a diphthong in the syllable, or (c) a vocalic element which ended the test word.

Responses To Be Considered for Analysis. To guarantee an adequate number of subjects for the observations of each test word, it was arbitrarily decided to eliminate from the analysis any word for which the test picture failed to elicit an appropriate response from at least 25 of the 30 subjects. This procedure guaranteed that there would be at least 25 responses for each test word, and thus the /r/ or /r/s in those test words, in the final analysis. In a few instances inconsistent pronunciation on the part of the subjects did not allow particular examples of /r/ to be classified consistently according to phonetic context. For example, some subjects pronounced *flower* as [flauwə˞] while others pronounced it as [flau˞]. The examples of /r/ in such words were also omitted from the data. Thus, the analysis is based upon 127 words; in all there were 144 examples of /r/ included for final analysis.

Reliability of Observers. The consistency of the observers in using pho-

TABLE 3. Percentages of correct articulations of /r/ within specific phonetic contexts. Entries in this table represent pooled data for 30 subjects.

Phonetic Context	*Percent of Correct Occurrences*	*Number of Observations*
[pr]	49.2	59
[dr]	40.3	119
[tr]	31.9	235
[str]	27.6	112
[br]	23.7	207
[ɪɚ]	23.7	59
[kr]	23.1	173
[ðɚ]	22.4	58
[gr]	22.2	90
[ɚt]	20.0	30
[ɚ]	18.6	414
[ɑɚ]	18.3	60
[skr]	17.9	29
[θr]	16.9	59
[fr]	16.8	89
[spr]	16.6	30
[r]	15.6	640
[ʃɚ]	14.8	27
[bɚ]	14.3	28
[pɚ]	13.5	89
[ɚm]	13.3	60
[ɛɚ]	12.8	149
[nɚ]	12.1	98
[lɚ]	11.7	60
[dɚ]	11.7	28
[sɚ]	11.1	27
[oʊɚ]	10.9	119
[mɚ]	10.3	58
[vɚ]	10.3	29
[ɑʊɚ]	10.3	29
[ɚz]	10.1	59
[ɚs]	10.0	30
[ɚd]	9.3	108
[tʃɚ]	9.0	89
[ɚ]	8.0	87
[tɚ]	7.3	108
[dʒɚ]	7.1	28
[gɚ]	7.0	115
[kɚ]	5.1	59
[ɑɪɚ]	5.1	59
[ɚk]	5.0	60
[ɚn]	3.4	88
[ŋɚ]	3.3	30

netic symbols to denote various /r/ responses (including deviations from normal /r/ articulations) was computed by dividing the total number of transcribed /r/ responses into the number of /r/ transcriptions which were identical for the two observers. The observers agreed on 86.6% of the responses transcribed. This percentage of agreement compares favorably with previously reported data (*2*, *3*, *4*, *6*, 7, *10*). It may be further noted that the observers' subjective evaluation of the method used to describe articulatory errors was that it was generally quite adequate. They were convinced also that such a system is generally more descriptive than a system using a broad category of 'distortion' to classify many errors of articulation.

Influence of Phonetic Context upon Misarticulation of /r/. In the data presented below, only those articulatory responses upon which both observers' transcriptions agreed exactly are indicated. All other /r/ responses were deleted. Thus, some tables will not show a total of 100% of the /r/ responses spoken. Table 3 presents the percentages of correct production of /r/ associated with each particular phonetic context for all 30 subjects.

The Glide Consonant [r]. These data for the glide consonant /r/ show, first, that in every instance the [r]s occurring in consonant blends were correctly articulated relatively more often than the single [r]s. Second, among the consonant blends the highest percentages of correct articulations of [r] occur for those blends which contain stop-consonants. Third, among the blends containing the stop-consonants, there seems to be a systematic variation of correct productions of [r] associated with the place of articulation of those consonants. Consonant blends containing front-stop consonants show higher percentages of correct articulations of [r] than do the blends containing back-stop consonants or continuant con-

TABLE 4. Percentages of subjects articulating at least one /r/ correctly within various phonetic contexts.

Phonetic Context	*Percent of Subjects*
/rV/	43.3
/Cr/	93.3
[ɚ]	66.7
/Vɚ/	60.0
/VɔC/	56.7
/Cɚ/	73.3
/CɔC/	16.7

sonants. The one exception is [spr] for which the percentage is based on a small number of responses since each subject had only one opportunity to articulate that combination.

The Unstressed Vocalic [ɚ]. Two points may be noted in the data for the unstressed vocalic /r/. (a) There seems to be much less variation in the frequency of correct production of the [ɚ] associated with any particular classification of context in which it occurred. However, as indicated for the glide [r], those contexts for [ɚ] containing sounds which may be classified as *front* sounds show higher percentages of correct productions than do those contexts containing back sounds; for example, [ɪɚ] and [ðɚ] may be compared to [kɚ] and [oʊɚ]. (b) The relative frequency of correct responses for the post-vocalic /r/ sounds is more similar to the other cases of the unstressed vowel than to the frequencies for the glide [r]. This feature of the data will be more fully discussed in a later section.

Comparisons among Subjects. Table 4 lists the percentages of subjects correctly articulating at least one /r/ within the different phonetic contexts. By comparing Table 4 with Table 3 it may be seen that the relationship between the number of subjects correctly articulating an /r/ within a given phonetic context and the percentage of correct productions of /r/ in that context is rather high. That is, a context in which /r/ occurs that was correctly articulated by a large percentage of subjects was also associated with a high percentage of correct productions of that /r/. This would certainly be expected if there were any consistency in responses from subject to subject. The fact that this relationship is marked indicates that the correct productions of /r/ in any context was not due to a few subjects but was a general characteristic of behavior of the group.

Differentiation among Types of /r/. Other factors specific to the articulation of /r/ may be pointed up from the data. The types of /r/, as defined in this study, can consistently be differen-

TABLE 5. Relative frequency of types of errors spoken for the various phonetic contexts. Table entries are percentages of all observations for each context.

	Type of Error										
Contexts	[w]	[$\overset{r}{w}$]	[əw]	*Omis.*	[ə]	[ɔ]	[ɔɔ]	[u]	[uə]	[ou]	[$\overset{o}{u}$]
/rV/	62.5	16.8		.3							
/Cr/	31.9	11.6	2.2	14.6							
[ɚ]					1.9	38.8	3.1	11.0	1.2	1.2	
/Vɚ/				10.9	41.0	1.9		10.7			
/VɔC/				33.0	38.4			2.0		2.7	1.1
/Cɚ/				.6	18.0	5.0		35.8		2.7	1.1
/CɔC/					5.5	17.2		4.5			

tiated on the basis of distinctive articulatory behavior. The data of Tables 3 and 4 show that, in general, the glide consonant [r] occurring in consonant blends ranked first in relative number of correct productions. Next, as ranked by percentage of correct production, is the stressed vocalic [ɝ]. Third is the unstressed vocalic [ɚ]. These differences in frequency of correct /r/ responses are generally consistent among types of /r/, irrespective of variations in phonetic context.

Table 5 points up the differential characteristics of these various types of /r/ by showing that the types of errors spoken for each type of /r/ are distinctive. Note that (a) the errors spoken for the consonant glide [r] were consonants; (b) the errors spoken for the stressed vocalic [ɝ] were vowels; and (c) the unstressed [ɚ] was like [ɝ] in that vowels always occurred as error responses, although, as mentioned above, it was unlike [ɝ] in the frequency of correct articulation.

Table 6, which gives the percentage of correct production of each type of /r/ by each subject, shows further evidence of the differential articulatory behavior with respect to types of /r/. Here it may be seen that the percentage of correct production of a given subject for one type, for example, consonant glide [r], provides little basis for predicting the frequency of correct production of the vowels [ɝ] or [ɚ] for that subject, and vice versa (compare Subject 1, who seldom produced a correct consonant /r/ but frequently articulated the vowel /r/s acceptably, with Subject 19, whose responses show the exact reverse).

Thus, the data of this study show several ways in which subjects behave differentially with respect to the different types of /r/, as here classified. It seems evident that this differential response should be given consideration when analyzing the problem of a child who exhibits misarticulations of /r/ sounds. Moreover, the evidence argues convincingly against classifying all of these /r/ sounds within a single category implying relationship with respect to articulatory movements. Rather, the evidence supports the division of /r/ sounds into a number of types to which responses seem to be sufficiently different to suggest that there may very well be important underlying differences of articulatory movement.

TABLE 6. Percentages of correct productions of each type of /r/ by subject.

Subject	/r/	/ɝ/	/ɚ/
1	3.5	84.6	59.4
2	8.6	7.7	1.6
3	8.0	.0	2.9
4	1.8	.0	.0
5	18.0	14.3	26.1
6	11.1	7.1	.0
7	6.3	35.7	40.6
8	50.8	7.1	2.9
9	40.6	14.3	5.7
10	58.1	78.6	16.4
11	.0	.0	2.9
12	23.8	.0	.0
13	1.6	.0	.0
14	62.5	50.0	13.0
15	14.1	28.6	8.6
16	65.6	.0	8.6
17	21.9	7.1	16.4
18	7.8	35.7	8.6
19	48.4	.0	4.3
20	6.3	.0	6.2
21	8.2	7.1	1.4
22	15.9	84.6	77.9
23	42.2	57.1	5.7
24	17.5	14.3	2.9
25	10.1	14.3	.0
26	31.0	46.2	7.2
27	1.6	.0	4.3
28	44.4	7.7	2.9
29	18.8	.0	.0
30	18.8	.0	1.4

TABLE 7. Percentages of correct articulations of /r/ in intersyllabic contexts. Selected data from Table 1 are included for comparison.

Intersyllabic Contexts			*Data from Table 1*		
Phonetic Context	*% of Correct Occurrences*	*Number of Observations*	*Phonetic Context*	*% of Correct Occurrences*	*Number of Observations*
[gɚ]	34.5	29	[gɚ]	7.0	115
[br]	33.7	89	[br]	23.7	207
[eɚ]	29.8	57	[eɚ]	12.8	149
[tɚ]	29.4	119	[tɚ]	7.3	108
[kr]	27.5	29	[kr]	23.1	173
[tr]	23.5	30	[tr]	31.9	235
[r]	18.6	113	[r]	15.6	640
[aɪɚ]	18.3	60	[aɪɚ]	5.1	59
[ɚr]	13.9	144			
[aɚ]	13.6	88	[aɚ]	18.3	60
[ɔɚ]	13.3	30			
[lɚ]	11.1	27	[lɚ]	11.7	60
[pr]	10.0	30	[pr]	49.2	59

The Postvocalic /r/. Another point of interest is that the data of this study support the point of view that the postvocalic /r/, as in board, park, car, functions as a vowel in the articulatory process. First, the relative frequency of correct production of the postvocalic /r/ in words such as those given was quite similar to that of the unstressed syllabic [ɚ] in water, mustard, ladder, which is universally accepted as being vocalic. Second, the sounds substituted for the postvocalic /r/ were always vowels, and were the same vowels in general as those substituted for [ɝ] and [ɚ] which are universally accepted as being vowels. It therefore seems clear that classification of the postvocalic /r/ as a vocalic element was justified, as was the use of the symbol [ɚ] to represent that phonetic element.

Intersyllabic Context. It has been previously mentioned that 29 of the /r/ examples included in the test were located between two syllables of the word so that an unambiguous decision could not be made assigning the /r/ to one or the other of the syllables. Since the assignment to syllables was uncertain, classification according to phonetic context was also equivocal. These cases will now be considered and compared to the data previously presented. In presenting these data in Table 7, the phonetic contexts were determined by the best judgment of the observers while listening to the transcription tapes. The uncertainty of this determination should, however, be borne in mind in considering these data. Selected data for comparable phonetic contexts from Table 3 are repeated in Table 7 so that comparisons may be readily made.

The number of observations of the intersyllabic /r/ contexts are small in six of the 12 instances. Those for which the number of observations listed in the table are 30, or less, occurred only once in the test, so that there is only one observation per subject. Thus, the percentage figures in those cases are inherently unstable. Despite this fact, some points may be made concerning these data: (a) Of the six contexts

which each subject had the opportunity to speak more than once, the higher percentages of correct /r/ productions are associated with front consonants and vowels. This is consistent with the previously presented data from Table 3. (b) There seems to be a general tendency in the direction of a greater frequency of correct productions of /r/s in intersyllabic contexts than in cases where the /r/ definitely initiates or terminates a syllable. This is particularly true of the unstressed vocalic [ɚ]. (c) Contrary to the previously presented data, there is little systematic ranking of percentages of correct production of /r/ by type of /r/.

This erratic behavior of /r/s in intersyllabic contexts as compared to the behavior of those in the main body of data may be due to the uncertainty with which phonetic context can be specified for these cases. Also, sounds occurring within a sequence of running syllables may vary sufficiently from similar sounds spoken as initiating or terminating elements that such differences may not be unexpected.

The Intervocalic /r/. In the group of /r/ responses occurring between connected syllables, one particular /r/ was differentiated from the other types of /r/. The observers realized early in the transcription procedure that the subjects' responses to the intervocalic /r/ of *arrow, carrot,* and similar words, were different from their responses to the other types of /r/. In particular, the types of misarticulation were distinctive. The dominant error with respect to this /r/ was that of omission. Each subject had five opportunities to produce this particular /r/ and for 42.6% of their attempts the omission error was noted. Furthermore, this /r/ was spoken correctly at least once by only 16 of the 30 subjects (56.7%). When the other types of /r/ are ranked according to the number of subjects speaking them correctly at least once, the intervocalic /r/ ranks last. The glide consonant [r] was spoken correctly at least once by 29 of the subjects, [ɝ] by 24 of the subjects, and [ɚ] by 20 of the subjects.

In the relatively few cases where the subjects substituted other sounds for the intervocalic /r/, the pattern of the error sound also was distinctive. The substitution was usually a vowel followed by a consonant and these elements were blended in such a way as to sound quite different from the errors noted with respect to the other types of /r/. Accordingly, the observers used the symbol [ɚr] to represent this intervocalic /r/ when it was correctly articulated.

In view of the above, it appears that there are at least three, and possibly four, types of /r/ sounds which may be sufficiently different in their characteristics to merit being considered as separate kinds of articulatory events, namely, (a) the consonant glide [r], (b) the stressed vocalic [ɝ], (c) the unstressed vocalic [ɚ], and (d) the intervocalic [ɚr].

Comparison with Previous Research. The previous research which is most comparable to the present study is the thesis of Buck (*2*). Previous discussion has pointed up certain differences in procedure and classification of sounds in the two studies. However, the data of the two are comparable in certain respects. With respect to the consonant glide [r] certain differences in the two

experimental groups' behavior may be noted: (a) For certain consonant blends, notably [pr], considerably higher percentages of correct production of the consonant [r] were observed in the present group than were noted for Buck's subjects. (b) Buck's subjects correctly articulated the glide [r] as a single relatively more often than they correctly articulated the glide [r] in certain blends, whereas, the subjects of this study correctly articulated the single consonant [r] relatively less than the [r] in any consonant blend. (c) The percentages of correct responses for /r/ in various contexts seem higher generally in this study than in Buck's. These differences may possibly be explained by the fact that Buck's subjects were restricted to three responses for each context. The subjects of the present study had more opportunities to speak the [r] in blends, for example, four times for [dr], seven times for [br], and six times for [kr]. The subjects of the present investigation had 22 opportunities to speak the single glide [r], which Buck's subjects articulated only three times. Thus the chances for correct articulations to occur are increased in the present instance.

Certain similarities in the data of the two studies also merit attention: (a) the postvocalic [ɚ] and postconsonant vowel [ɚ] showed similar percentages of correct production in both studies. (b) In both studies the stressed vocalic [ɝ] showed higher percentages of correct production than did the unstressed vocalic [ɚ]. (c) In both studies relationships appeared between the percentages of correct /r/ production in consonant blends and the tongue movement or placement of the sound blended with /r/. Buck noted that occurrence of blends with consonants requiring definite tongue position or movement, especially postdentals, seem to facilitate correct [r] productions. Subjects in the present study correctly produced [r] more frequently when it was blended with front-stop consonants.

More generally, the data of the present study certainly lend support to the findings of previous workers (*2, 3, 10, 11*) that correct production of consonant sounds usually misarticulated as singles may be facilitated in blends. The present results also corroborate previous researches (*1, 2, 3, 10, 11, 15*) in that inconsistent misarticulation seems to be the rule rather than the exception in the speech patterns of children with functional articulation problems.

Discussion

The findings of this study point to several general influences which are of sufficient significance to be re-emphasized.

a. On the basis of the articulatory behavior analyzed in this study it would appear that the consonant glide [r], the stressed vocalic [ɝ], the unstressed vocalic [ɚ], and, perhaps, the type of /r/ heard in arrow ([ɚr]), are different classes of phonetic events to which subjects show differential behavior. Grouping them into a general classification for clinical or analytical purposes ignores these distinctions. It may be further suggested that the differences in response patterns shown in this study suggest underlying differences in articulatory movements which may need to be considered in relation to the problem of their appropriate phonemic classification.

b. The postvocalic [ɚ] is best considered as a vocalic element since articulatory responses in words containing this /r/ are similar to other vowel /r/ sounds and dissimilar to consonant /r/ responses.

c. This study is in agreement with previous research in finding that inconsistency of misarticulation is common in the speech of children with functional misarticulation. Data of this study and of Buck's investigation indicate that with respect to the various types of /r/ this inconsistency is systematic and can be related to a certain extent to the phonetic context in which the /r/ occurs.

d. These data, as do the data of previous studies dealing with inconsistent misarticulation, indicate that a rather thorough phonetic analysis of articulatory defective speech may be valuable in understanding the problem and planning a corrective program. A detailed analysis would seem indicated both to determine the specific types of sounds which are misarticulated and to reveal phonetic contexts which facilitate correct articulation for the individual case.

e. The data of this investigation point strongly to the conclusion that phonetic context is an important factor in the articulatory process and suggests that there may be a need to re-examine the common assumption that clinical retraining procedures should usually begin with isolated sounds on the presumption that they are more readily articulated.

Summary

The consistency of misarticulations of the /r/ phoneme was analyzed for a sample of 30 children who had been selected because of functional difficulty with /r/. A system of narrow phonetic transcription was used to describe the misarticulations observed on magnetic recordings of a word articulation test containing 175 examples of /r/.

The results show characteristic differences in response patterns among different consonant /r/s, among certain subclasses of vowel /r/s, and among different phonetic contexts. Types of /r/ could be differentiated with respect to both the frequency and the types of error responses. Characteristically different patterns of error responses were obtained for vowel /r/s as compared to consonant /r/s.

Comparison is made with previous data from similar research and clinical implications are suggested. The data also suggest that certain /r/ occurrences which have been frequently classified as consonant articulations may be more appropriately considered to be vowels.

Acknowledgment

The authors are indebted to Dr. Don Dew of the University of Maryland who, with Mr. Hardy, spent many hours phonetically transcribing the subjects' production of the test words.

References

1. Amidon, Hilda F., A statistical study of relationships among articulatory errors made by one hundred first grade children, M.A. thesis, Univ. of Iowa, 1941.
2. Buck, McK. W., A study of the misarticulation of [r] in children from kindergarten through third grade, M.A. thesis, Univ. of Iowa, 1948.
3. Hale, Anita R., A study of the misarticulation of [s] in children from kindergarten through third grade, M.A. thesis, Univ. of Iowa, 1948.

4. Henderson, F. H., Accuracy in testing the articulation of speech sounds. *J. educ. Res.*, 31, 1938, 348-356.
5. Hultzen, L. S., Symbol for the nonsyllabic post-vocalic R of General American. *Quart. J. Speech*, 34, 1950, 189-201.
6. Irwin, O. C., and Chen, H. P., A reliability study of speech sounds observed in the crying of newborn infants. *Child Develpm.*, 12, 1941, 351-368.
7. Irwin, O. C., and Curry, T., Vowel elements in the crying vocalization of infants under ten days of age. *Child Develpm.*, 12, 1941, 351-368.
8. Kenyon, J. S., and Knott, T. A., *A Pronouncing Dictionary of American English*, Springfield, Mass.: G. & C. Merriam Co., 1949.
9. Moses, E. R., Jr., Palatography and speech improvement. *J. Speech Dis.*, 4, 1939, 103-114.
10. Nelson, J. T., A study of misarticulation of [s] in combination with selected vowels and consonants, M.A. thesis, Univ. of Iowa, 1945.
11. Roe, V., and Milisen, R., The effect of maturation upon defective articulation in elementary grades. *J. Speech Dis.*, 7, 1942, 37-50.
12. Spriestersbach, D. C., and Curtis, J. F., Misarticulation and discrimination of speech sounds. *Quart. J. Speech*, 37, 1951, 483-491.
13. Stetson, R. H., *Motor Phonetics*. Amsterdam: North-Holland Pub. Co., 1951.
14. Thomas, K. C., *Phonetics of American English*. New York: Ronald Press, 1947.
15. Wilson, Lois, A survey of speech defects among mentally retarded pupils in the public schools of Des Moines, Iowa, M.A. thesis, Univ. of Iowa, 1945.

Training Parents of Children with Functional Misarticulation

RONALD K. SOMMERS

SAMUEL P. SHILLING

CLARA D. PAUL

FLORENCE G. COPETAS

DOLORES C. BOWSER

COLETTE J. McCLINTOCK

Nature of the role of the parent in a speech therapy program is a matter of concern to most speech therapists. The degree to which a parent should participate in a rehabilitation program for cerebral palsy speech (*2*), stuttering (*3*, pp. 225-226), and cleft palate speech (*8*, p. 606) has been discussed by some writers.

The parental role in articulation therapy also has been studied and in the area of articulation problems there are divergent opinions on this role. Powers (*8*, pp. 800-801) feels that the 'cooperative nature of articulation therapy' suggests that parents should engage actively at an appropriate point in therapy in assisting the therapist with the correction of the problem. West, Ansberry, and Carr (*10*, pp. 352-353) state that parents 'should be given considerable responsibility in the program.' Lillywhite (*6*) feels that optimum results will be obtained in the correction of speech defects if mother is made a therapist. On the other hand, Wood (*11*) found that mothers of children with functional articulation problems were more neurotic in tendency than mothers of 'normal' children. He concluded that mothers of children with functional articulation problems should be dissuaded from undertaking speech correction procedures with the child in his home.

The present study was designed to evaluate the results of training parents to help at home children with functional articulation defects.

Ronald K. Sommers (M.A., Northwestern, 1953) is Supervisor of Speech and Hearing, Armstrong County, Pennsylvania. Samuel P. Shilling (M.S., Pennsylvania State University, 1954) is Director of Speech Therapy, New Castle Public Schools, Pennsylvania. Clara D. Paul (M.Ed., Pennsylvania State University, 1955), Florence G. Copetas (M.Litt., University of Pittsburgh, 1950), Dolores C. Bowser (B.S., California State Teachers College, Pennsylvania, 1954), Colette J. McClintock (M.Ed., Pennsylvania State University, 1958), are Speech Therapists, Armstrong County, Pennsylvania. This article is based on a paper presented at the American Speech and Hearing Association 1958 convention, New York.

Procedure

Subjects. In the experimental group were 36 children, 10 girls and 26 boys; in the control group were 36 children, nine girls and 27 boys. All children in both groups had functional articulation problems and all were Caucasian except for a Negro boy in the experimental group. They were enrolled in two clinics (Armstrong County, Pennsylvania, 1957) conducted simultaneously 20 miles apart, the experimental group at one clinic and the control group at the other.

At the time of enrollment in each clinic and after completion of appropriate oral examinations of teeth alignment and spacing, degree of dental occlusion, tongue mobility, height of palatal arch, and velar movement, all children who evidenced functional articulation problems were further tested by a staff of six speech therapists. They were given the Otis Quick Scoring Group Intelligence Test, Forms A and B, to estimate mental ages. Their hearing was screened at 20 db from 250 through 8000 cps using Beltone 10A audiometers. Their speech was tested by the use of a combination of flash cards drawn from the Developmental Articulation Test and the Warnock-Medline Cards to meet the five criteria established by Wright (*12*) for reliable testing on the following consonant sounds used in this study: [r], [l], [s], [f], [v], [k], [g], [tʃ], [ʃ], and [θ]. Pictures of objects used on these cards were those selected by Wright except for the addition of those for [θ] and [ʃ]. Each subject was tested by each therapist on all 10 sounds, each sound appearing in the initial, medial, and final positions, in the order reported above. Criteria for judging responses, directions to the subjects, and methods of recording responses all followed the procedures established as reliable by Wright (*12*). Errors of omission, substitution, and distortion, as well as correct responses, were recorded.

A subject was not used in this study if he had a hearing loss of 20 db or greater at any frequency; if his mental age, as determined by the Otis tests was less than five years and 11 months; or if his oral structure was judged to be anomalous in any way.

The method of equivalent groups described by Garrett (*4*, pp. 211-212) was the basis for equating the experimental and control groups. The groups of subjects were matched on the basis on four variables: sex. I.Q., number of days of clinic attendance, and number of articulation errors. The effectiveness of the matching procedures may be judged by noting the approximate equality for the two groups of the means and standard deviations on preliminary measures. For the experimental and control groups the mean chronological ages were seven years, four months and seven years, seven months with standard deviations of 1.05 and 1.41 years, respectively; mean mental ages were eight years, one month and seven years, five months with standard deviations of 1.32 and 1.54; and corresponding mean I.Q.'s were 102 and 98 with standard deviations of 4.81 and 3.98. Mean times for previous therapy were nine months, four days and eight months, five days, with standard deviations of 0.73 and 0.45 months, for the experimental and control groups, respectively. Mean numbers of articulation errors were

5.72 and 6.50, with standard deviations of 2.55 and 2.83 errors, for the experimental and control groups, respectively.

Articulation Tests. The experiment involved articulation tests at three stages: Pre-, Post- I, and Post- II. The Pretesting was done immediately before the 3.5-week clinical period. Post- I testing was done directly after the clinical period. Post- II testing was done nine weeks after Post- I.

Each articulation test at all three stages involved identical procedures and flash cards in testing for omissions, substitutions, and distortions on the 10 consonant sounds chosen for the study. Post- I testing of the experimental group was by a therapist from the control group and, on the same day, Post- I testing of the children in the control group was done by a therapist from the experimental group. The same arrangements were made for Post- II testing so that in both Post- I and Post- II testing no therapist tested any child he knew or had diagnosed or treated during the clinical period.

Reliability of Therapists. The reliability of the six therapists who served as judges in the experiment was tested in two ways: (a) The percentage of agreement among the six testers was determined when two live subjects were tested, one therapist presenting the flash cards and all therapists judging the responses simultaneously. The mean percentage of agreement for all raters was 87.7%. Mean agreement on omissions was 89.6%; on substitutions, 88.0%; on distortions, 83.1%; and on correct sounds, 90.7%. (b) During the Pretest the same flash cards and procedures were used to test the 10 consonant sounds in three positions. Each rater compared his agreement with every other rater and himself on the same subject. The intragroup mean agreement was 86.7%: that is, 86.7% of the time the six raters agreed whether an omission, a substitution, a distortion, or a correct sound was produced. No attempt was made to refine this mean percentage of agreement into mean agreement on omissions, substitutions, distortions, and correct sounds.

Therapy. Speech therapy was provided by a staff of three experienced speech therapists in each clinic. A chief therapist in each clinic administered the program and consulted with the parents in varying degrees, as noted later. Each clinic was in session during the same hours of the day; classes were one hour in length. Therapy was provided four consecutive days weekly in each clinic for a period of three and one-half weeks.

The articulation therapy used for both groups might be classified generally as the phonetic-placement approach advocated by Van Riper (*9*, pp. 175-184) and described by Berry and Eisenson (*1*, pp. 163-164). In addition, there were games and activities similar to those described by Van Riper (*9*) and also his four steps for ear training: isolation, stimulation, identification, and discrimination. Certain aspects of the sound-stimulation method of articulation therapy as discussed by Berry and Eisenson (*1*, pp. 162-163) and Johnson and others (*5*, pp. 126-129) supplemented the phonetic-placement method. Mirrors, audio and visual aids, and the previously mentioned notebooks were utilized by therapists in both groups.

Therapists in each clinic were instructed and reminded throughout the clinical period to emphasize as many of the 10 consonant sounds as children in a therapy group required. Each child was given speech therapy on all of his faulty sounds. For some children this meant all 10 sounds.

Three weekly meetings of the entire staff of both clinics were held during the course of the clinical period to increase the homogeneity of therapy techniques. During these the techniques being used by each therapist were carefully reviewed.

In both clinics children were assigned for therapy in groups on the basis of both their mental ages and the consonant sounds with which they had problems. The mean number of children in a speech therapy class in the experimental group was 5.8 and in the control group, 5.5.

Work with Parents. At the beginning of speech therapy for the control group, the parents were told simply to attend when possible. For the experimental group, however, the chief therapist indicated to the parents that their own attendance at the clinic was important and necessary. This therapist developed a program for the experimental-group parents which was presented at the same time that their children attended therapy classes. There were daily 30-minute lecture discussion periods when such topics as these were considered: What is a speech problem? How do children develop speech? What materials are needed to help children learn to correct speech problems? What is ear training and what are the techniques of ear training? What are some placement techniques and how are they used? How can the child be stimulated to better speech production at home without pressuring or nagging him? What devices for ear training can be developed that will work at home to accomplish speech correction without accompanying tension?

The experimental-group parents observed speech therapy after the discussion period each day. In the course of the program each parent observed two different speech therapists working with separate groups of children and using techniques which included ear training and phonetic placement. Then on the day following such observation, the parents discussed with the chief therapist what had been observed and how comparable incidents might be handled at home.

The speech therapists who worked with the children in the experimental group were encouraged, in addition, to meet with the parents for a short time after each session to give specific advice and daily home assignments, the latter being written by the therapist for each child in a speech notebook. Furthermore, at the end of the clinical program the therapists listed for the parents of each child the defective consonant phonemes requiring more therapy and suggested some techniques that could be used during the remainder of the summer vacation. The parents of children in the experimental group were told that their children's articulation would be evaluated again in September (three months later).

The procedure for parents of children in the control group was markedly different. No parent here observed speech therapy with his child at any

time. No specific instructions were provided directly to the parents by the therapists concerning sounds to be stressed at home, the nature of a particular child's speech problem, suggestions for home help. Each child was provided with a speech notebook and was given assignments by his therapist but the therapist did not explain these assignments to the parents of the child. If a parent persisted in wanting this information, his child was eliminated from the study.

Parents of children in the control group met with the chief speech therapist on three occasions. The chief therapist avoided talking about techniques of helping the child either on specific sounds or in general; he gave no indication that such an effort on the part of the parents was expected or desired. Instead, general topics were discussed: how children develop speech, what articulation is, and how teasing affects children with speech problems. All meetings were brief and controlled by the therapist. If a parent inquired about his child or about some specific aspect of therapy, a cursory and general remark was made avoiding specific suggestions for parental participation at home. After three sessions these parental meetings were discontinued in the control group. No outline for parental help during the remainder of the summer was provided nor was any reference made to the fact that their children's articulation would be evaluated again in September.

Questionnaire. At the end of the clinical period the parents of both groups, anonymously and by free choice, answered a questionnaire giving their own and their children's attitudes toward the clinic and their estimates of the benefit their children received from attending. Items were scaled by the parents on a continuum from *least* to *most*. For example, for the question, 'Did you gain in understanding of your child's speech problem?' the parent chose one of the following: gained nothing, gained a little, gained some, gained a lot, or gained greatly.

Results

A total of 30 articulation errors was possible for each subject since each of the 10 consonant sounds was tested in initial, medial, and final positions. In the Pretest 206 articulation errors were found in the experimental group (parental training) and 234 in the control group (no parental training). Of the 206 errors for the experimental group, 136 (66%) were substitutions, 37 (18%) were omissions, and 33 (16%) were distortions. Of the 234 for the control group, 145 (62%) were substitutions, 39 (16.6%) were omissions, and 50 (21.4%) were distortions. Thus, in both groups the substitution type of error was found to be approximately three times as frequent as either omissions or distortions. It is to be remembered, however, that this analysis is for the total number of errors without reference to individual subjects. This finding agrees with the Roe and Milisen (7) report on the incidence of types of articulation errors in children at these age levels.

Mean differences between the experimental and the control groups with respect to changes in number of errors between any two of the articulation tests (Pre-, Post- I, Post- II) were eval-

uated by the *t* test for related measures. Separate tests were made for correctly articulated sounds, for substitutions, for omissions, and for distortions. Only four of the 12 obtained *t* values were significant. The mean reduction of number of substitution errors between the Pretest and Post- I test was 1.05 for the experimental group and .25 for the control group, with the difference significant at the 2% level (t=2.52 and df=29; N varies from one test to another depending upon the number of subjects with the type of error tested). The corresponding means for the Pretest and Post- II test were 1.86 and 1.22, with the reduction of substitution errors significantly greater, at the 5% level (t=2.21 and df=29), for the experimental group than for the control group. The mean reduction of numbers of omission errors between the Post- I test and the Post- II test was .90 for the experimental group and 4.70 for the control group; this difference is significant at the 5% level (t=2.50 and df=22). The reduction of frequency of distorted sounds between the Post- I test and the Post- II test was 1.15 for the experimental group in contrast to an increase of .64 for the control group. This difference is significant at the 1% level (t=2.94 and df=27). Most of the remaining eight differences between the groups are quite small and all are nonsignificant.

Both groups of parents indicated by their answers to the questionnaire generally favorable attitudes toward the therapy program. There was, however, some tendency for the parents of the experimental group to answer all questions slightly more positively than the control group parents, although differences between the groups were not statistically significant.

Discussion. The data in this study reveal a trend which suggests that the simultaneous training of parents and children with functional articulation problems may result in more rapid improvement of articulation than would be the case if the parents received no training. Improvement under these conditions was most evident in the substitution type of articulation error, on which the experimental group made significantly greater improvement than the control group, an improvement made during the clinical period and maintained throughout the entire period of the study. This result is particularly interesting since more than 60% of the articulation errors made by each group were of the substitution type. Neither group made noticeable progress during the clinical period in correcting distortions. In fact, the number of distortions slightly increased during this period for the experimental group. After the clinical period there was a like increase for the control group. These increases are perhaps to be expected in view of the findings by Roe and Milisen (*8*), whose study points out that substitutions decrease with age and distortions increase with age. Parent training may have little bearing on reduction of consonant sound omissions since neither group improved more than the other in this respect during the total period of the study. The control group, however, did show more improvement than the experimental group on omissions during the postclinical period, a significant difference for which there is no readily apparent explanation.

Although some statistically significant differences between groups were found, the magnitude of the absolute differences may not appear to be of clinical importance. It is suggested that had the period of the clinics been longer, these differences might have assumed greater clinical importance and that parent training over a longer span of time might have resulted in more improvement in articulation than was the case with this brief period of intensive training.

The results of this study do not provide evidence to support Wood's (*11*) feeling that mothers of children with functional articulation problems should be dissuaded from helping them at home. It should be mentioned, however, that the approach in this study might not be duplicated in many home situations. In this study care was taken to develop pleasant, indirect, and anxiety-free suggestions on home assignments for parents of the experimental group. Furthermore, drill work was given to very few children and then only to selected ones who were judged to be sufficiently mature mentally and emotionally for this kind of therapy.

The responses to the questionnaire were very similar for the two groups of parents, although somewhat more positive for the experimental group than for the control. This lack of difference may to some extent be implicit in questionnaire studies; also it may indicate that parents who enroll their children in such a program are a select group who tend to show appreciation and gratitude.

Summary

Two groups of children, 36 in each group, with functional articulation problems were matched on the basis of four factors: chronological age, mental age, sex, and number of misarticulated sounds. Parents of children in the experimental group received intensive training in techniques of helping their children at home. Parents of children in the control group received no training.

Articulation tests were administered just prior to the 3.5-week clinical program, immediately following it, and again nine weeks later. The two groups showed about the same improvement in increasing the mean number of correct sounds. The experimental group showed more improvement by reduction of the number of substitutions and distortions than did the control group. The control group showed a postclinic-period reduction of omissions significantly greater than that of the experimental group. Parent attitudes toward the clinic program and their children's improvement, as indicated in questionnaire answers, were not significantly different for the two groups.

Acknowledgment

Appreciation is hereby extended to Dr. George Shames of the University of Pittsburgh for his encouragement and consultation services and also to Dr. Eugene McDonald of Pennsylvania State University for his interest in this project.

References

1. Berry, Mildred F., and Eisenson, J., *Speech Disorders*, New York: Appleton-Century-Crofts, 1956.
2. Coffman, Ruth, and Backus, O., Group therapy with preschool children having cerebral palsy. *J. Speech Hearing Dis.*, 18, 1953, 350-354.

3. Eisenson, J., and Ogilvie, Mardel, *Speech Correction in the Schools*. New York: MacMillan, 1957.
4. Garrett, H. E., *Statistics in Psychology and Education*. New York: Longmans, Green, 1947.
5. Johnson, W., Brown, S. J., Curtis, J. F., Edney, C. W., and Keaster, Jacqueline, *Speech Handicapped School Children* (rev. ed.). New York: Harper, 1956.
6. Lillywhite, H., Make mother a clinician. *J. Speech Hearing Dis.*, 13, 1948, 61-66.
7. Roe, Vivian, and Milisen, R., The effect of maturation upon defective articulation in elementary grades. *J. Speech Dis.*, 7, 1942, 37-50.
8. Travis, L. E. (Ed.), *Handbook of Speech Pathology*. New York: Appleton-Century-Crofts, 1957.
9. Van Riper, C., *Speech Correction: Principles and Methods* (3rd ed.). New York: Prentice-Hall, 1954.
10. West, R., Ansberry, M., and Carr, Anna, *The Rehabilitation of Speech* (3rd ed.). New York: Harper, 1957.
11. Wood, K. S., Parental maladjustment and functional articulatory defects in children. *J. Speech Dis.*, 11, 1946, 255-275.
12. Wright, H. N., Reliability of evaluations during basic articulation and stimulation testing. *J. Speech Hearing Dis.*, Monogr. Suppl. 4, 1954, 19-27.

Sidetone Delay and Reading Rate, Articulation, and Pitch

G. DON DAVIDSON

In the communication process the human organism makes use of an auditory system by which he monitors various characteristics of his vocalization. This hearing of his own speech is commonly designated by the term 'sidetone.' This auditory perception is not simultaneous with vocalization. Both the distance from the vocal cords to the cochlea and the type of medium through which the sound must travel effect a delay in the arrival of the oral signal at the cochlea.

In recent years there has been some interest in electronically retarding or accelerating the normal external sidetone delay and then measuring the effect of the new transmission time on various aspects of speech. After vocalizing under a condition of sidetone retarded more than the normal amount, it has been the subjective experience of this writer and others that a sudden switch to the normal delay seemed to facilitate the speaking process. Going from a normal delay condition to an accelerated one seemed to produce the same sensation but with a more pronounced effect.

The possible facilitation of speech under accelerated sidetone suggests certain questions. Is there an accompanying change in the precision of articulation? Is there a change in pitch variability? Is there a change in the number of inflections? Is there a change in the extent of inflections? Is there a change in the mean rate of inflection?

It was the purpose of this study to determine the effect of the external sidetone transmission times of approximately 0.0005, 0.0010, and 0.0015 sec upon the following: (a) the oral reading rate of normal speakers; (b) the precision of articulation; and (c) pitch variability as measured by the number, the mean extent, and the mean rate of inflections.

Procedure

Apparatus. Three microphones (condenser type, Altec-Lansing, one Model 21-B and two Model 21-C with associated P518-A power supplies) served in the delay unit. A fourth microphone (cardioid type, Western Electric, Model 639) picked up the subject's speech for recording. A three-channel mixer was constructed with the use of three T-pads (Mallory, 250 ohm, 4 watt); three single-pole, double-throw switches; and appropriate resistors. The mixer served to produce a level signal output from all three power supplies. The output of the three-channel mixer was sent to an amplifier (Presto, Model 85-B). This

G. Don Davidson (Ph.D., Ohio State University, 1955) is Assistant Professor of Audiology and Speech Pathology, Florida State University. This article is based upon a doctoral dissertation completed under the direction of Dr. John W. Black and Dr. Dorothy Sherman.

amplifier in turn fed the signal to the subject's headphones. A tape recorder (Magnecorder, Model PT6-P) served to provide a recording of the subject's vocal responses. An attenuator (Hewlett-Packard, Model 350-A, 6 watt, 500 ohm) provided a regulation of the signal voltage across the subject's headphones. A voltmeter (Hewlett-Packard, Model 400-C) served to measure the voltage across the subject's headset and to determine the adjustment of the attenuator. The subject received his sidetone through a pair of headphones (Model ANB-H-1A). The experimental sessions were conducted in a room that had been padded on the walls, ceiling, and floor to minimize reverberations of the subject's speech which might have modified the delay experience at his ears.

Determination of Delay Times. Three sidetone delay times were used in the experiment: 0.0005, 0.0010, and 0.0015 sec. The delay time of 0.0010 sec approximated the average normal sidetone delay time (*3*). This time was obtained in the experiment by placing one of the condenser microphones 6 in. from and directly in front of the speaker's mouth. A little under 6 in. represents the average distance from the lips to the tragus. Since the transmission time through an electronic system is negligible in this case, placing the microphone 6 in. from the mouth approximates the normal delay time. Moving the microphone in either direction from this point would produce a delay time longer or shorter than normal. The longest delay time, 0.0015 sec, was obtained by a second microphone placed 12 in. from and directly in front of the speaker's mouth. The shortest delay time, 0.0005 sec, was obtained by placing a third microphone at the right corner of the speaker's mouth. The delay times of 0.0005 and 0.0015 sec are approximations based on the assumption of a normal delay time of 0.0010 sec and on the fact that sound travels through the air at about 1000 feet or 12000 in. per sec or 6 in. in 0.0005 sec.

Stimuli. The stimulus material consisted of five five-syllable phrases selected from the Walker-Black phrases (*4*), a factual 70-word reading passage, and Speaker Lists 1 through 24 of Forms C and D of the POS Multiple-Choice Intelligibility Test (*1*).

Subjects. The subjects were 24 college-age males with normal hearing. None had had previous experience with abnormal sidetone delay.

Experimental Procedure. Each of the 24 subjects read the five five-syllable phrases, the reading passage, and the five five-syllable phrases again under each of the three sidetone conditions. (The five five-syllable phrases were read a second time for another purpose beyond the scope of this article.) Order of sidetone conditions was counterbalanced. It was assumed, however, for purposes of analysis, that practice would have eliminated any order effect. In addition, eight of the 24 subjects each read three Speaker Lists of the multiple-choice intelligibility test, one for each sidetone condition. The order of the sidetone conditions was independently randomized for each of the eight subjects. Means for conditions would thus not be biased by order effect. Responses were recorded on tape at a speed of 7.5 in. per second.

Immediately preceding the first experimental condition, each subject practiced reading the stimulus material aloud according to the procedure to be

used later. The subject was instructed that each of the five five-syllable phrases was to be read without pause or hesitation, with a two-second interval between adjacent phrases as signaled by the experimenter. The interval between the phrases and the reading passage was three seconds. The practice procedure was repeated three times. Eight of the subjects then practiced reading their intelligibility lists three times. They were instructed to read each word series within a list without hesitation and as if it were a meaningful phrase. A two-second interval came between adjacent word series within a list and approximately three to five seconds between lists. During the practice period any errors in pronunciation were corrected. Also during the practice period a constant vocal level for any of the stimulus material was to be monitored on the VU meter of the tape recorder. This level was approximately 82 db (reference 0.0002 dyne/cm^2) as measured six inches from the mouth on a General Radio sound level meter on the C scale. By adjusting the T-pads of the three-channel mixer it was possible to provide at the headphones a signal voltage constant from one delay condition to another if the subject spoke at equal voice levels under all three microphone conditions. During the experiment, because of possible distraction from the experimental task, the subjects did not monitor vocal level. The experimenter observed, however, that all subjects tended during the experiment to maintain approximately the vocal level used during the practice period.

Criterion Measures. The criterion measure of oral reading rate was the mean duration of five five-syllable phrases in seconds determined from power-level recorder tracings measured in millimeters and converted to seconds.

The criterion measure of precision of articulation was the intelligibility score on the multiple-choice intelligibility test. Tape recordings of the Speaker Lists were edited and arranged to allow use of the standard scoring sheet. These recordings were then presented in 70 db of noise (S/N ratio –6 db to –8 db, reference 0.0002 dyne/cm^2) to 23 listeners. An intelligibility score was determined for the panel of eight speakers under each delay condition for each listener.

The three criterion measures of pitch variability were the number, the mean extent of inflections, and the mean rate of inflections in a selection from the reading passage 'You might think it easy to figure out each new idea, but changes come slowly.' This particular sentence was chosen because of its relatively great pitch variability and because it was neither at the beginning nor at the end of the passage. The sentence was recorded at a speed of 30 in. per second and then played at 3.5 in. per second. The output of the tape recorder was fed to an Edin Oscillograph Pen Writer. The resulting tracings were analyzed for number of cycles per .04 sec. Frequencies were converted to semitones on the musical scale (2) and the number, the mean extent, and the mean rate of inflection determined.

Results

Oral Reading Rate. Differences among the three delay conditions and also between the first and second readings were significant with respect to mean phrase duration in seconds for the read-

TABLE 1. Summary of variance for evaluation of sidetone delays and two readings with respect to duration of reading.

Source	*df*	*ms*	*F**	$F_{.05}$
Delay Time (D)	2	.0210	11.11	3.20
Reading (R)	1	.0111	8.54	4.28
Subjects (S)	23	.0956		
DR	2	.0023	1.77	3.20
DS	46	.0018		
RS	23	.0013		
DRS	46	.0013		
Total	143			

*F ratios: ms_D/ms_{DS}; ms_R/ms_{RS}; ms_{DR}/ms_{DRS}.

ing of the five five-syllable phrases by 24 speakers (see Table 1). Mean duration[1] decreased with progressively shorter delay times. Differences between the shortest delay time mean and the two longer delay time means were each statistically significant.[2] The mean duration[3] of the second reading was significantly less than that of the first. Consideration of the difference between readings, as previously mentioned, is not relevant to the purposes of the present discussion. Of importance, however, is the fact that, in view of the nonsignificant interaction between readings and delay times, the averages of the two readings may be used for the comparisons among delay times.

Precision of Articulation. Differences among the three delay conditions on mean intelligibility scores, obtained for a panel of eight speakers for each of 23 listeners, were nonsignificant.[4] The evidence thus does not support the hypothesis that precision of articulation varies with changes in rate resulting from altering sidetone-delay time. Another test than the one employed in this experiment (POS Multiple-Choice Intelligibility) might, however, result in more precise measures and provide evidence to support such a hypothesis.

Pitch Variability. Differences among the three conditions of sidetone delay for number of inflections used by 24 speakers in reading of the factual 70-word passage were nonsignificant.[5] The effect of the sidetone conditions upon mean extent of inflections expressed in semitones was, however, highly significant.[6] The mean extent for the shorter-than-normal sidetone-delay time was

[1]The observed means were 1.322, 1.347, and 1.363 for the 0.0005, 0.0010, and 0.0015 sec delay times, respectively.

[2]Required difference (r.d.) for significance $= t(2ms_{error}/n)^{1/2}$ r.d. $_{.05} = .0168$; r.d. $_{.01} = .022$.

[3]The observed means were 1.353 and 1.335 for the first and second readings, respectively.

[4]$F = ms_D/ms_{DS}(df = 2 \text{ and } 44) = 13.5/13.6 < 1.00$.

[5]$F = ms_D/ms_{DS}$ $(df = 2 \text{ and } 46) = .39/2.91 < 1.00$.

[6]$F = ms_D/ms_{DS} = .505/.079 = 6.39$; $F_{.01}$ (df = 2 and 46) = 5.10.

significantly[7] greater than for the other two delay times. The effect of the sidetone conditions upon mean rate of inflection in semitones per second was also highly significant.[8] Again the significant[9] differences were between the shorter-than-normal sidetone-delay time and each of the other two delay times, with the greatest rate of inflection for the shorter-than-normal delay time. The results thus provide some evidence to support the previously mentioned hypothesis that the shorter-than-normal delay time facilitates the speaking process. The assumption seems reasonable that such facilitation is indicated by the significant increases in extent of inflection and rate of inflection, as well as by a significant increase of reading rate, which accompany the shorter-than-normal delay time.

Summary

The purpose was to evaluate the effect upon oral reading of three sidetone transmission times: shorter-than-normal or 0.0005 sec, normal or 0.0010 sec, and longer-than-normal or 0.0015 sec. The measures taken were (a) reading rate, (b) precision of articulation, and (c) three measures of pitch variability: the number, the mean extent, and the mean rate of inflection.

Observed reading time increased with each increase in delay time; differences between the shorter-than-normal and each of the other delay times were significant. Extent and rate of inflection were significantly greatest for the shorter-than-normal delay time. There was no evidence that the other variables studied were affected by the changes in delay times.

References

1. Black, J. W., Multiple-choice intelligibility tests. *J. Speech Hearing Dis.*, 22, 1957, 213-235.
2. Moon, P. A., A scale for specifying frequency levels in octaves and semitones. *J. acous. Soc. Amer.*, 25, 1953, 506-515.
3. Stromsta, C., A first approximation of the distance from vocal cords to cochlea and the transit time of bone-conducted sound from the region of the vocal cords to the region of the cochlea. M.A. thesis, Ohio State Univ., 1951.
4. Walker, C., and Black, J .W., The intrinsic intensity of oral phrases. U. S. Naval School Aviation Medicine, Pensacola, Bur. Med. Res. Proj. NM 001 053, *Joint Report No. 2*, 1950.

[7] $r.d._{.05} = .159$; $r.d._{.01} = .209$. Observed means = 2.363, 2.077, and 2.179 for delay times of 0.0005, 0.0010, and 0.0015 sec, respectively.

[8] $ms_D/ms_{DS} = 92.09/8.80 = 10.46$;
$F_{.01}$ (df = 2 and 46) = 5.10.

[9] $r.d._{.05} = 1.44$; $r.d._{.01} = 1.89$. Observed means = 22.29, 18.55, and 19.42 semitones per second for delay times of 0.0005, 0.0010, and 0.0015 sec, respectively.

Lipreading Performance and the Rate of the Speaker

VINCENT W. BYERS

LEWIS LIEBERMAN

In an effort to understand better the process of lip reading, several studies have been conducted in which the relationship between lip-reading 'ability' as a trait and other psychological variables have been investigated. Lip reading and such factors as intelligence, perceptual skills, educational achievement and personal adjustment are discussed by Heider and Heider (*3*), O'Neill (*5*), and Pintner (7). O'Neill and Davidson (*6*) have studied concept formation, reading comprehension and level of aspiration.

The present study is an investigation of the relationship between lip-reading performance and the speaking rate of the person whose lips are being read, speaking rate being a variable more intrinsic to the lip-reading process than many of the other factors that have been studied previously. With the aid of a motion picture camera, it is possible to photograph a person speaking and to project this photograph at different speeds. Since the measurement of speech rate (words per minute) is a rather nonprecise measurement, the motion picture camera introduces a more precise control of speaking rate than is likely by merely instructing a person to speak slower or faster. Also changing a speech rate involves a consideration of the relationship of phonation time and pause time. Use of different projection speeds of a camera produces different rates of speech but preserves the original relationship of phonation time to pause time. Since speech at 'normal' rate is already quite rapid, the rates chosen for this particular study were slower-than-normal. Since reduction in rate might produce different effects relative to the abilities (calibre) of the lip reader, the present study divided the subjects, all experienced lip readers, into a 'good' and a 'poor' group.

Procedure

Subjects. The subjects were 94 males and females, all students at the Kentucky School for the Deaf, Danville, Kentucky. Each had had at least two years of formal lip-reading training. Since all were required to write simple

Vincent W. Byers (M. A., University of Kentucky, 1956) is a graduate student, University of Pittsburgh. Lewis Lieberman (M.S., University of Kentucky, 1958) is a graduate student, University of Kentucky. This article is adapted from a paper presented at the 1958 Convention of the American Speech and Hearing Association in New York. The study was supported in part by the Department of Psychology, University of Kentucky.

sentences, a minimum age of 11 years was set.

Materials. As measures of lip-reading performance, the two forms of the Utley sentence test for lip reading (*8*) were used. The sentence tests were selected because of their reported high reliability (*2*, *8*) and high validity (*2*). The tests were filmed in order to equate stimuli for subjects tested on different occasions and to insure control of speaker rate. Restrictions in amount of available film made it necessary to reduce the number of items in the Utley tests. Form A was reduced from 31 sentences with 125 words to 24 sentences with 91 words. This reduction was accomplished with the aid of data reported by DiCarlo and Kataja (*2*): those sentences with the greatest number of words which correlated poorly with the test as a whole were eliminated until there were 24 left. Sentences eliminated were numbers 2, 11, 20, 24, 26, 29, and 30. This adaptation of Utley's test will be referred to here as the Experimental Film. Form B also was reduced. From 30 sentences with 126 words, 24 sentences with 86 words were retained. Since no correlations were available, eliminations were made on an arbitrary basis, such as length, stilted English, and insufficient difficulty. The following sentences were eliminated: numbers 3, 19, 24, 25, 26, and 29. This adaptation of Form B will be referred to as the Screening Film. The Screening Film included Utley's five practice sentences which were not used in scoring the subjects' performances.

There were three parts to the film, all using the same speaker. The speaker was a young female, trained in speech and dramatics, who was judged to have general American dialect and superior speech. She was instructed to speak each sentence with the same 'expressionless' countenance in order to cut down extraneous visual cues. The camera was focused to include only the head and shoulders of the speaker. Make-up, including lip rouge, was not used. These measures were taken on the advice of DiCarlo and Kataja (*2*), who criticized the Utley films for the distraction caused by the arm movements and the make-up of the speaker.

The three parts of the film were (a) the Screening Film photographed at normal speed (24 frames per second); (b) the Experimental Film photographed at normal speed; and (c) the Experimental Film photographed in 'slow motion' (48 frames per second). The format for each part was the same: a title announcing the sentence number was flashed on the screen for a second; this was followed by a clip of the speaker reciting the sentence; this was followed by 15 seconds of blank film during which the subjects could write down the sentence just presented. Care was taken that the speaker would be seen briefly both before she spoke and at the completion of each item to prevent a choppy, distracting presentation.

The silent films were presented to the subjects on a Bell and Howell 16 mm projector with two speeds. Each film could thus be shown at 16 frames per second or 24 frames per second. The films were shown in a semidarkened classroom in which a maximum of 14 subjects could be conveniently accommodated.

Instructions. Before each session, printed instructions were given to the subjects. To make sure the subjects comprehended them, the instructions were reviewed both verbally and manually by one of the Danville staff. The instructions served four purposes: (a) to motivate the subjects by telling them that this was part of a study to improve movies for lip readers, (b) to avoid creating a test situation, (c) to explain the format, and (d) to encourage subjects to guess and to write down whatever parts of the sentences they understood.

Experimental Procedure. The entire sample of 94 subjects was first tested with the Screening Film shown at normal speed. Subjects wrote their responses on specially prepared answer sheets. Tests were scored according to Utley's procedure, counting one point for each correct word.

On the basis of the Screening Film, the 24 subjects with the highest scores and the 24 with the lowest scores were selected for the experiment proper which was conducted two days later. The members of each subgroup of 24 were randomly assigned to one of four experimental groups. In this way, each experimental group consisted of 12 subjects, six good lip readers and six poor lip readers. The experimental conditions were as follows: Group I (control), tested on the normal Experimental Film projected at normal speed with a speaker rate of 120 words per minute; Group II, tested on the normal Experimental Film projected at 16 frames per second with a speaker rate two-thirds of normal or 80 words per minute; Group III, tested on the slow motion Experimental Film projected at normal speed with a speaker rate one-half of normal or 60 words per minute: and Group IV, tested on the slow motion Experimental Film projected at 16 frames per second with a speaker rate one-third of normal or 40 words per minute.

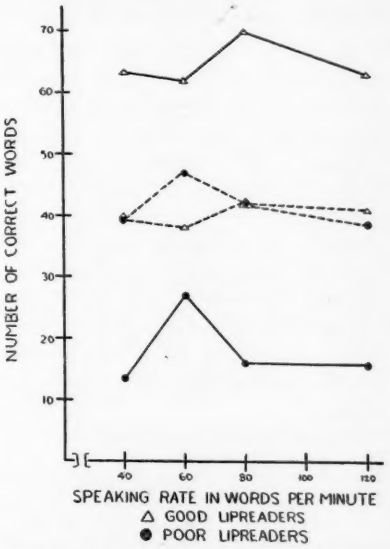

FIGURE 1. Group means for the good and poor lip readers for the number of correct words at each of the four speech rate conditions. The dotted lines connect the adjusted means of the analysis of covariance with the screening film scores as the predictor variable.

In view of the lack of precision of the unit of rate (words per minute), no elaborate operations were adopted to measure the actual elapsed speaking time. The reported rates were calculated from measures taken with a stop watch during runnings of the film. Rates were then rounded off to the nearest five words per minute for simplicity of reporting.

Results

Plotted in Figure 1 are the group means for the good and poor lip readers at the four rates. For the good lip readers, the group means are as follows: Condition I, 63.33 correct words; Condition II, 69.50 correct words; Condition III, 62.00 correct words; and Condition IV, 63.17 correct words. For the poor lip readers, the group means are as follows: Condition I, 15.67 correct words; Condition II, 15.67 correct words; Condition III, 26.67 correct words; and Condition IV, 13.00 correct words.

TABLE 1. Analysis of variance of correct responses.

Source	*df*	*ms*	*F*
Calibres (C)	1	26226.75	269.90*
Rates (R)	3	97.25	1.00
C R	3	193.03	1.99
Within cells (error)	40	97.14	
Total	47		

*Significant beyond .001 level.

The most deviant mean for the good group occurred at 80 words per minute; the most deviant mean for the poor group occurred at 60 words per minute. To test whether there were significant differences among the means, an analysis of variance (Table 1) was employed, with calibre (good-poor), rates, and interaction (calibre x rates) as between-group variables. The only significant result was the difference between the means of the good and the poor (calibre), which is to be expected in view of the way the good and poor groups were selected.

It is possible that, in the random assignment of subjects to experimental conditions, true differences were obscured by chance. To investigate this possibility an analysis of covariance was employed with the Screening Film scores as the 'predictor' variable. The rates and interaction *F* ratios were still not significant. The adjusted means are plotted in Figure 1 (dotted lines) and it can be seen that the apparent discrepancies are no longer as marked. As expected, the *F* ratio for calibre was not significant in the covariance analysis.

Although there were no significant differences among means for the four rates with respect to the number of correct responses, a further consideration is the total number of responses which subjects made. Results of an analysis of variance based on the total number of intelligible words each subject produced, however, were similar to results of the analysis for correct words, with calibre being the only significant effect.

In summary, the results suggest that over the range studied, the rate variable is not significant in lip-reading performance for either good or poor lip readers, either in terms of quantity of words produced or in number of correct responses.

Discussion

Although very little has been written about the rate variable per se in the lip-reading process, the feeling seems to be, for example, as stated by Nitchie (*4*, p. 18), that on the continuum of possible rates, 'normal' speech rate is faster than optimum lip-reading rate, that is, lip reading would be easier if people's normal rate were slower. Unless this optimum is below 40 words per minute,

the results of this study appear to contradict this expectation.

On the other hand, a safe conjecture would be that at both ends of the rate continuum, at a point too fast and at a point too slow, lip reading would be impossible. In the section of the continuum studied here, there was no noticeable decline in group performance, in spite of the fact that one of the rates was one-third of the normal rate and in spite of several complaints from the subjects that the slow motion films were too slow.

The admonition (*1*) given to instructors of lip reading not to slow down their rates is, of course, recognition of the fact that lip reading is usually done in a situation where speaking rate is normal, but this same admonition might suggest also that lip-reading skill is not adaptable to changes in rate. The present study suggests that the skill is adaptable to quite an extensive range of clearly distinguishable rates. Before such a generalization can be made, however, one must know that this adaptability is reversible, that is, that individuals taught at slower rates would be just as efficient at normal rates. There are no experimental data, yet, to settle this question.

The lack of evidence that slowing of rate affects the performances of either the good or the poor lip readers also seems to contradict the expectation of certain workers in the field. Nitchie (*4*) seems to feel that one of the factors which produces a good lip reader is a 'quick mind,' which he goes on to describe as 'a mind which is quick to respond to impressions or quick in its reaction time' (*4*, p. 19). It would follow from this that the poorer lip reader would improve if it were possible to increase his quickness. If one grants the equivalence between slowing down rate of stimuli and increasing rate of response, this prediction is not substantiated by the present results. It is possible that the element of naturalness of face movements, disturbed when the film was slowed, might have had some effect on the findings. Although this was not within the scope of the present study, it might be a proper subject for further research.

Summary

Four groups of experienced lip readers were chosen from students of the Kentucky School for the Deaf. Each group consisted of six 'good' and six 'poor' lip readers, separated on the basis of a filmed lip-reading test and randomly assigned to the four groups. Each group was exposed to a filmed version of a sentence lip-reading test (adapted from Utley); the first group viewed the film at 'normal' rate (120 words per minute), the second group saw the film at two-thirds that rate, the third at one-half that rate, and the fourth group at one-third that rate. Results showed no significant differences among the four rates either in number of words correct or in quantity of words produced for either the 'good' or 'poor' lip readers. Neither was there any significant interaction effect between calibre (good-poor) and rates.

Acknowledgments

Photography was done by Richard G. Sanderson of the Radio Arts Department, University of Kentucky. The authors gratefully acknowledge the help of Dr. James S. Calvin, University

of Kentucky; Dr. Madison J. Lee, superintendent, Miss Isabelle Walker, academic principal, and the staff and students of the Kentucky School for the Deaf.

References

1. Bunger, Anna Mae, *Speech Reading—Jena Method*. Danville, Ill.: Interstate, 1952.
2. DiCarlo, L. M., and Kataja, R., An analysis of the Utley lipreading test. *J. Speech Hearing Dis.*, 16, 1951, 226-240.
3. Heider, F. K., and Heider, Grace M., Studies in the psychology of the deaf. *Psychol. Monogr.*, 52 (1), 1940, 124-153.
4. Nitchie, E. B., *Lip-reading Principles and Practice*, rev. by Elizabeth H. Nitchie and Gertrude Torrey. New York: Frederick A. Stokes, 1930.
5. O'Neill, J., An exploratory investigation of lipreading ability among normal hearing students. *Speech Monogr.*, 18, 1951, 309-311.
6. O'Neill, J. J., and Davidson, JoAnn L., Relationship between lipreading ability and five psychological factors. *J. Speech Hearing Dis.*, 21, 1956, 478-481.
7. Pintner, R., Speech and speech-reading tests for the deaf. *J. appl. Psychol.*, 13, 1929, 220-225.
8. Utley, Jean, A test of lipreading ability. *J. Speech Hearing Dis.*, 11, 1946, 109-116.

Speech Pickup by Contact Microphone at Head and Neck Positions

JOHN C. SNIDECOR

IRVING REHMAN

DAVID D. WASHBURN

In certain military operations it would be desirable to pick up speech by a contact microphone from a source other than the front of the mouth for the following reasons: (a) in aircraft the microphone would not interfere with oxygen equipment; (b) the valve pops and breathing hisses of the oxygen equipment would not affect speech intelligibility; (c) the possibility of injuries to the mouth and teeth would be minimized when the pilot bails out of an aircraft at high speed; (d) in high noise levels, such as those in the vicinity of jet aircraft during ground operations, communicators would gain some advantage from observing lip movement; (e) certain pickup locations are better for noise shielding; for example, an ear canal microphone can be shielded by a helmet or by muffs and a helmet, and a contact microphone on the forehead or mandible by a helmet.

One disadvantage of such application is apparent: vibratory speech pickup from these locations suffers some distortions and attenuation when compared to speech picked up in the air in front of the mouth.

Recent investigations by Mullendore (*9*), Békésy (*2*), Moser, Dreher, and Oyer (*8*), and University of Iowa researchers (*1*) show that pickup of vibratory energy during vocal production varies over a considerable range according to the pickup site on the head, neck, and other upper body parts.

The Békésy data (*2*) and the data of Moser, Dreher, and Oyer (*8*) indicate that, in general, as distance from the larynx increases, pickup intensity decreases. There are, however, reversals in this trend. In the Iowa data (*1*), pickup intensities at the sternum, elbow, and shoulder were relatively high.

The present preliminary study is part of a larger program oriented toward the selection of microphones and pickup sites which give satisfactory in-

John C. Snidecor (Ph.D., University of Iowa, 1940) is Professor of Speech, University of California, Santa Barbara. Irving Rehman (Ph.D., New York University, 1939) is Associate Professor of Anatomy, Medical School, University of Southern California, and Consultant, U. S. Navy Electronics Laboratory and San Diego Naval Hospital. David D. Washburn (B.S., Purdue University, 1950) is Psychologist, Auditory Detection and Communications Section, Human Factors Division, U. S. Navy Electronics Laboratory. This study was completed at the Navy Electronics Laboratory, Human Factors Division, San Diego.

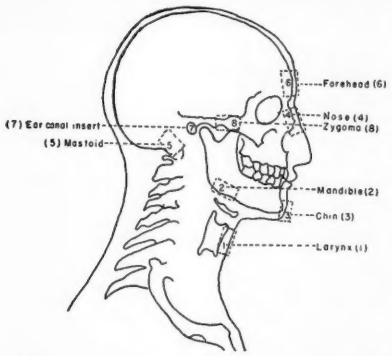

FIGURE 1. Anatomical locations used for vibratory speech pickup with a contact (crystal) microphone.

telligibility, quality, and usability under high ambient noise conditions. Divided into two parts, the present study investigates (a) the relative power of the vowels [i] (meet), [ɛ] (met), [ɔ] (loss), and [u] (room); and (b) the relative quality preference for a standard sample of continuous speech (*4*) as picked up by a crystal vibration microphone at eight different locations on the head and neck.

Procedure and Results

Relative Power of Vowels. A Brush VM-1 Vibromike (crystal) was used as the contact microphone. Magnetic tape recordings were made of speech picked up from each of eight locations on the head and neck of an experienced talker (see Figure 1). At each location, five recordings were made for each of four vowels, [i], [ɛ], [ɔ], and [u]. These five trials were made by removing and replacing the contact microphone five times during continued phonation of each of the four vowels. A short speech sample also was recorded at each position.

The talker monitored his phonation at 82 db on the C (flat) scale of a General Radio sound level meter, with its microphone approximately 14 inches in front of his lips. Simultaneous magnetic tape recordings were made of the output from each of these microphones (sound level meter and VM-1). Comparisons of Bruel and Kjaer power-level tracings from both recordings permitted correction of data for variations in vocal effort. Speech-power data were based upon these corrected Bruel and Kjaer tracings. The speech-sample recordings were used to judge intelligibility and pleasantness of quality in the second part of the experiment.

A three-way analysis of variance (four vowels, eight locations, and five

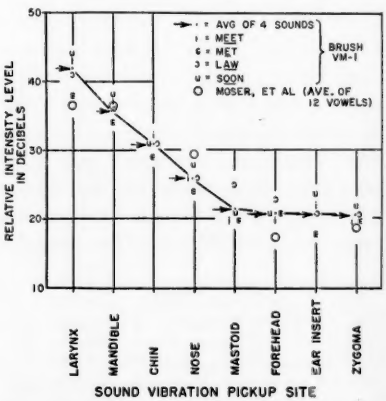

FIGURE 2. Mean levels (five trials each) for each of the four vowels, [i], [ɛ], [ɔ], and [u], as well as the overall means for all four vowels (connected points with arrows) picked up by contact microphone at each of eight head and neck locations. Open circles indicate averages of data for all vowels at comparable locations of the study by Moser and others (7).

trials) was run to determine whether the differences among vowels and among locations were significant with variance among trials used as the error term. Figure 2 shows the mean (over five trials) level for each of the four vowels for each pickup location. The differences among locations are significant at the 0.1% level while the differences among vowels are significant at the 1% level. The open circles in Figure 2 are the averages from the Moser and others (7) study and agree quite well with the results of the present study in regard to speech power among locations. The results of this study show that the vowel [u] has the highest average power while [ɛ] has the least. The means plotted in Figure 2 indicate that each of eight locations supplies enough energy to be potentially useful.

Quality Preference. The authors judged the intelligibility of the recorded speech samples picked up by the contact microphone, basing their judgments on a five-point scale (*poor*, *fair*, *average*, *good*, and *excellent*). They judged the speech picked up from the mandible, chin, nose, mastoid process, forehead, and ear-canal insert to be *good;* and from the larynx (thyroid cartilage) and zygoma to be *average*. With regard to the larynx, this conclusion has been reached by other experimenters doing research on throat microphones. Since intelligibility was judged to be approximately equal from six of the locations, quality was evaluated to select the locations presenting speech of the most pleasing quality.

The authors were under the impression that three of the locations giving *good* intelligibility might scale quite differently on listener preference rating. First, the forehead gave relatively clear, crisp speech with good consonant articulation. Second, the mandible gave speech easy to understand, but not especially pleasant in quality. Finally, the nose bridge (nasion) gave *good* intelligibility, but also had a nasal, strident quality and loud breath sounds.

TABLE 1. Paired comparison preferences of 24 judges for quality of speech picked up at eight positions on one talker; also chi square test scale results.

Position	*Total Preferences*	*Scale Value*
Forehead	140	1.340
Mastoid	93	.445
Larynx	93	.445
Seventh cervical vertebra	90	.407
Zygomatic arch	70	.324
Temporo-mandible joint	69	.122
Mandible	63	.042
Nose	50	.000

A brief sample of speech from each of the eight positions was compared once with a similar sample from each of the other positions by the method of (A-B) paired comparison and forced preference choice. Twenty-four advanced college students majoring in speech correction and phonetics served as judges. They were instructed to state a preference for one speech sample from each of the 28 pairs of samples.

The scaling of these judgments was computed by a method developed by Guilford (*6*), and the results are shown in Table 1.[1] Speech from the forehead,

[1]Statistical computations on paired comparisons and scaling were done by Dr. Francis P. Leukel of San Diego State College, San Diego, California.

the mandible, and the nose were favored in that order when only the three locations were considered. All three were significantly different, at the 1% level ($df = 2$), from the preference pooled for the other seven judgments. Differences among the three were not statistically evaluated because of the *post hoc* logic involved, which disallows such statistical treatment. Chi square for differences in preference as compared to a random distribution among the eight locations is 63, which is statistically significant at the 1% level ($df = 7$).

It is clear that speech from the forehead was very much preferred by the group, while speech pickup from the nose was the least preferred. It should be emphasized that in service situations listener preference is secondary to intelligibility.

Conclusions

On the basis of power, judged intelligibility, and listener preference, one can assume that the forehead, mastoid, and larynx are favored for additional study. On the basis of power and potential intelligibility, the mandible, ear canal, and nose merit further study even though the qualities from the nose and mandible are less preferred. Lower power and preference probably exclude the zygomatic arch.

Final acceptance of instruments and points of pickup depend not only upon intelligibility and quality in quiet, but also upon other factors, including intelligibility in noise, convenience of placement, and comfort. Research in progress indicates that the Brush VM-1 may be placed in such a way that it is comfortable and convenient but low in noise discrimination. Since it was not designed for the application herein studied, no criticism is implied in terms of the functions for which it was designed. Intelligibility tests in noise appear desirable for some of the locations explored. Contact microphones, either crystal or dynamic, may soon be available which will better meet service criteria for comfort, convenience of placement, and intelligibility in high-level ambient noise.

Summary

Preliminary investigation based on one talker was made into the intensity, intelligibility, and quality preference of speech recorded from a contact transducer (vibration microphone) at various head locations. From each location, the vowels [i], [ɛ], [ɔ], and [u] and a short speech sample were recorded on magnetic tape, from which sound-level tracings were made. Twenty-four judges rated the recorded speech by A-B paired comparisons for pleasing quality. Intelligibility was judged by the authors. The following locations give promise of being suitable pickup positions during military duties requiring that no lip or free-field microphone be used: forehead, mastoid process, larynx, mandibular angle, ear canal, and nose.

References

1. Atkinson, C., Research study of the psychoacoustic effects of human artificial sidetone. Fifth Quart. Rep. Sig. Corps Supply Agency and Univ. of Iowa, Contract DA 36-039-SC-42562, 1953.

2. Békésy, G. v., and Rosenblith, W., The mechanical properties of the ear. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley, 1951.
3. Black, J. D., An ear-insert microphone. *RCA Engr Dig.*, Jan.-Feb., 1956, 5-10.
4. Fairbanks, G., *Voice and Articulation Drillbook*. New York: Harper, 1940.
5. Fletcher, H., *Speech and Hearing*. New York: Van Nostrand, 1929.
6. Guilford, J. P., *Psychometric Methods*. New York: McGraw-Hill, 1936.
7. Moser, H., Dreher, J., and Oyer, H., The relative intelligibility of speech recorded simultaneously at ear and mouth. AF Cambridge Res. Center TN 55-64, Ohio State Univ. Res. Fndn. TR-28, May 1956.
8. Moser, H., Dreher, J., and Oyer, H., The relative intensity of sound at various anatomical locations on the head and neck during phonation of vocal sounds. AF Cambridge Res. Ctr. TN 55-70, Ohio State Univ. Res. Fndn. TR-31, June 1956, PB 124-137.
9. Mullendore, J., An experimental study of the vibration of the bones of the head and chest during sustained vowels. *Speech Monogr.*, 16, 1949, 163-176.

Voice Quality and Anxiety

CHARLES F. DIEHL

RICHARD WHITE

KENNETH W. BURK

Despite a considerable mass of literature in the area of voice and personality (*1*, *6*) only three studies, known by the authors, have investigated the relationship between voice quality and personality traits. Duncan (*1*), using a sample of college students, compared scores on the Bell Adjustment Inventory with judgments of vocal efficiency and found positive correlations between low social adjustment and inefficient use of quality, pitch, and force. The most significant and stable correlations were between voice quality and low social adjustment. In a further study, Duncan (*2*) compared Bell Adjustment Inventory scores of 22 persons having functional hoarseness with those of a control group free of vocal anomalies. Ten subjects in the hoarse group differed importantly from the control group in home adjustment scores.

Moore (*5*), using the Bernreuter Personality Inventory, compared 43 college students with superior speaking voices and 119 college students whose voices were classified variously as breathy, harsh, metallic, and whiny. Analysis of results indicated that (a) the speaker with a breathy voice is likely to be high in neurotic tendency and introversion, (b) the speaker with a harsh and metallic voice tends to be dominant and emotionally stable, and (c) the speaker with nasal whine tends to be emotionally unstable and low in dominance although somewhat more dominant than the breathy speaker.

The data reported in these studies seem generally to provide evidence of a relationship between voice quality and personality adjustments. However, since many of the correlations are low, and measuring instruments used may have questionable validity, any conclusions based on these data can be considered only as tentative. On the other hand, there appears to be no evidence which tends to refute a hypothesis associating voice quality with personality manifestations.

The present article reports the results of an investigation in which an answer was sought to the following question: What is the relation of voice quality to anxiety as measured by the Taylor anxiety scale? It was assumed

Charles F. Diehl, (Ph.D., Pennsylvania State University, 1949) is Professor of Psychology and Director of the Speech Center, University of Kentucky. Richard White, (Ph.D., University of Kentucky, 1958) is Fellow in Speech, College of the Bible, Lexington, and Research Assistant, University of Kentucky Speech Center. Kenneth W. Burk, (M.A., University of Kentucky, 1955) is Instructor in Psychology and Speech Therapist at the University of Kentucky Speech Center.

for the purposes of this study that the several voice quality deviations to be considered could be identified as clinical entities.

Anxiety, defined as a 'free-floating anticipatory fear reaction,' (7), was measured with the Taylor personality scale of manifest anxiety. On the basis of construction and reported performance, the Taylor scale (4) appears to have at least fair validity for certain purposes: that is, the anxiety scale scores are significantly related to clinical judgments of anxiety in psychiatric cases. Also, according to three studies by Taylor (7), the scale has a test-retest reliability of .89, .82, and .81, respectively.

Procedure

Subjects were 169 males and 10 females, with a mean age of 25 years, selected randomly from college students enrolled in two seminaries, one in Lexington, and one in Louisville, Kentucky. The subjects came from 28 states. States with largest representation were: Kentucky, 41; Ohio, 20; Virginia, 16; Illinois, 13; California, Tennessee, and Alabama, each nine. The remaining 51 subjects were about equally distributed among the remaining 21 states.

Each subject was asked to read the same passage from the Bible, Matthew 5, which was chosen because it was believed to be familiar to all subjects. Each subject was instructed to read the passage as if he were in the pulpit Sunday morning and no subject was permitted to study the passage prior to the reading. All readings were recorded with the same equipment by the same experimenter in approximately the same situation (a small room with no other individuals present). Three of the subjects who demonstrated severe hoarseness during the voice recordings were referred to an otolaryngologist for examination. The medical reports for these three subjects were negative. Subjects who admitted having a cold, sore throat, or respiratory disorder of any kind, were asked to reschedule their recordings for a later date when they were well.

The Taylor scale was completed by all subjects either singly or in groups after the recordings were made. All subjects were assigned a code number for both recording and test, and were assured that they would remain anonymous.

The recordings were judged by the three authors in the following manner:

During the first of two listening sessions, all recordings were judged for 'normal' or 'defective' voice quality. Voice quality was labeled 'defective' if in the opinion of each judge it deviated sufficiently from 'normal' to warrant clinical attention. In all cases where judges could not reach unanimous agreement, the subjects were eliminated from the sample. Recognizing the possibility of certain regional characteristics of speech among the various subjects, the judges attempted to control this variable. They agreed in advance, for example, that a voice would be called defective in nasality only if it evidenced more nasality than that which could be attributed to regionalism. Judges were not familiar with the anxiety scores of the subjects during the listening sessions.

The second judgments were made

TABLE 1. Summary of analysis of variance for testing differences among voice quality groups (normal, hoarse-breathy, nasal, and harsh) with respect to mean anxiety scores.

Source	*df*	*ms*	*F*	$F_{.01}$
Between groups	3	237.91	4.44	3.89
Within groups	171	53.57		
Total	174			

only on those subjects classified as 'defective.' Each judge was instructed to place each subject into one of three categories of predominant defectiveness: harsh, nasal, or hoarse-breathy. Fairbanks' (*3*) definitions of four basic voice quality types (harsh, nasal, breathy, hoarse) were used as a basis for these classifications. His hoarseness and breathiness categories were combined into one on the basis of a pilot study in which the three experimenters concluded that hoarseness and breathiness, in their opinions, had certain similar acoustical characteristics and could be assigned by them with greater validity to a single group (hoarse-breathy) than to two groups (hoarse, breathy).

Results

Of the 178 subjects, 113 were judged 'normal' and 62 'defective' in voice quality. Three subjects, about whom the judges could not agree, were eliminated from the sample.

The 62 subjects judged as having defective voice quality were categorized into the following groups: harsh, 19; nasal, 18; and hoarse-breathy, 25. None of the subjects with a voice deviation was eliminated for lack of agreement as to predominant type of defectiveness.

According to analysis of variance of mean anxiety scores of four groups (normal, harsh, nasal, hoarse-breathy), differences among groups were significant beyond the 1% level (see Table 1). The group means and results of the significance of differences among group means are given in Table 2. The mean of the anxiety scores for the hoarse-breathy group is significantly greater than the mean for the normal group and also the mean for the harsh group. Other comparisons between groups were not significant.

Conclusions and Discussion

Any evaluation of the data must take into account the artificial and circumscribed character of the testing situation, the limited nature of the sample, and the special meanings assigned to 'normal' and 'defective' as applied to speech qualities. The results, in general and within those boundaries, support those reported by Duncan and Moore. The assumption must be made, however, in comparing Duncan's results to

TABLE 2. Results of *t* tests between mean anxiety scores* on the Taylor scale of manifest anxiety; groups compared: normal (1), hoarse-breathy (2), harsh (3), and nasal (4).

Groups	*N*	*Mean*	*S.E.diff.*	*t*
1	113	12.64	1.74	3.31†
2	25	18.40		
1	113	12.64	1.63	.07
3	19	12.53		
1	113	12.64	1.95	.90
4	18	14.39		
2	25	18.40	2.19	2.68**
3	19	12.53		
2	25	18.40	2.44	1.64
4	18	14.39		
3	19	12.53	2.36	.79
4	18	14.39		

*The 50th percentile on the Taylor = 13.
†Significant at the 0.1% level.
**Significant at the 2% level.

the present ones, that poor social and home adjustment (which she found correlated positively with inefficient use of voice quality and hoarseness, respectively) are related to anxiety states.

The evidence from Moore's study (in which he found positive correlations between breathiness and neurotic tendencies and introversion, and between harshness and dominance and emotional stability) is supported by the present results. On the basis of Moore's data it is reasonable to predict more anxiety for a hoarse-breathy group than for a normal or a harsh group. Since such a prediction is substantiated by the results reported here, a positive relationship between hoarseness-breathiness and anxiety, and a lack of relationship between harshness and anxiety seem more defensible than previously.

There is no immediate explanation for failure to find a relationship between nasality and anxiety in this study, a result which might be predicted on the basis of Moore's conclusion that a speaker with nasal whine tends to be emotionally unstable and low in dominance although somewhat more dominant than the breathy speaker. Some part of the reason for this failure might be found in possible differences between 'nasality' as used in this study and 'nasal whine' as used in Moore's study, a voice quality for which he gives no definition.

The shifting circumstances under which all individuals speak, the interrelationship of all the voice variables during spontaneous speech, and the lack of appropriate instruments with which to isolate and measure both voice and personality variables under realistic conditions, make definitive research in this area difficult. There seems evidence to justify further research into the relationship between voice quality and personality.

Summary

The purpose was to investigate relationships between voice quality and anxiety as measured by a personality scale of manifest anxiety (Taylor A). The subjects, 178 seminary students, read orally a passage from the Bible. They were instructed to read as if in the pulpit Sunday morning. The voices were judged and classified under the following voice qualities: normal, harsh, nasal, hoarse-breathy. The four groups were compared on the basis of scores on the anxiety scale. The results indicated, within the limits imposed by the experimental situation, that persons with hoarse-breathy voices are more anxious than (a) persons with normal voices and (b) persons with harsh voices.

References

1. Duncan, Melba H., An experimental study of some of the relationships between voice and personality among students of speech. *Speech Monogr.*, 12, 1945, 47-60.
2. Duncan, Melba H., Personality adjustment techniques in voice therapy. *J. Speech Dis.*, 12, 1947, 161-167.
3. Fairbanks, G., *Voice and Articulation Drillbook*. New York: Harper, 1940.
4. Lauterbach, C. G., The Taylor A scale and clinical measures of anxiety. *J. cons. Psychol.*, 22, 1958, 314.
5. Moore, W. E., Personality traits and voice quality deficiencies. *J. Speech Dis.*, 4, 1939, 33-36.
6. Sanford, F. H., Speech and personality. *Psychol. Bull.*, 39, 1942, 811-845.
7. Taylor, Janet, A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 48, 1953, 285-290.

Experimental Blockage of Phonation by Distorted Sidetone

COURTNEY STROMSTA

The relative importance of feedback systems to the neuromuscular synergy necessary for speech has been theorized, at least since the advent of cybernetics (*33*). Research concerning one such system, termed sidetone and constituted by the hearing of one's own speech via bilateral bone-tissue and air-conduction pathways, has demonstrated a close relationship between auditory perception and speech. Black (*8*) and Lee (*19, 20*), utilizing a 0.18- to 0.20-sec delay of air-conduction sidetone, observed increased intensity, slower rate, and a disturbance to speech similar to stuttering. On the other hand, acceleration of air-conduction sidetone was observed by Peters (*25*) to facilitate the rate of speaking. Dolch and Schubert (*11*) have reported that when accelerated air-conduction sidetone was transmitted to the ears 180 degrees out-of-phase to the signal emitted at the mouth, the subject experienced harshness of voice and spoke with greater effort, at a slower rate, and with increased intensity.

Courtney Stromsta (Ph.D., Ohio State University, 1956) is Assistant Professor of Speech, Ohio State University. A portion of the article was presented at the 1958 Convention of the American Speech and Hearing Association, New York. The research was supported by a grant from the National Institutes of Health, Division of Neurological Diseases and Blindness.

Little information is available that would explain the underlying system being affected by delayed air-conduction sidetone. Electrophysiological results by Baccaro (*4*) indicated that cortical potentials were significantly greater in amplitude and lower in frequency for subjects reading under a condition of 0.20-sec delay as compared to no delay. Fairbanks and Guttman (*12*) theorized '. . . that disturbed articulation and increased duration are "direct effects" of the delay; that greater sound pressure and higher fundamental frequency are "indirect effects," evidencing "effort to maintain system control" and to "resist experimental interference with the response"; and that articulatory disturbance is the "primary effect". . . .'

Vocal cord activity as well as speech per se has been reported to have been affected by auditory perception. Vannier, Saumont, Labarraque, and Husson (*31*) studied the effects on phonation of shifting the phase of air-conduction sidetone, using a vowel sustained in falsetto as the signal. On the basis of stroboscopic examination they reported a rapid intermittent blockage of vocal cord activity within a phase interval, which, when translated into time, was of the order of 0.7 to 0.8 msec. The blockage reportedly occurred only

when the vowel was sustained in falsetto and at a low level in accompaniment with the altered sidetone transmitted to the ears with 'high amplification.'

The explanation offered for the blockage of vocal cord activity presupposed a cochleorecurrential reflex. Husson (*18*) conceived of the following pathway for the reflex arc: auditory nerve, cochlear nuclei, reticular formation, sensory and motor nuclei of the trigeminal, recurrent motor nuclei, and laryngeal musculature. As described by Vannier and others (*31*) a recurrent nerve synapse enters a refractory period lasting 0.7 to 0.8 msec subsequent to the arrival of an auditory nerve impulse and is therefore not excitable during that period. If a nerve impulse arrives at the synapse during that period, the departure of any recurrent nerve impulse is blocked.[1]

Implicit in the foregoing experimental results and ensuing theorizations would seem to be a system which, when its limiting conditions are exceeded, reflects itself peripherally as speech or voice variations. If it can be assumed that the controlling factor is a composite signal transduced by the basilar membranes to the central nervous system which in turn must integrate or correlate the transduced signals, then it is apparent that the aggregate effect of bilateral bone-tissue and air-conduction sidetone must be appreciated.

In this regard available theory and basic information allow only a first approximation as a frame of reference. Békésy (*5*) has stated that the experience of air-conduction sidetone can be considered as equal in magnitude to that of bone-tissue sidetone. Nonetheless there are reasons why their similarity can be questioned in terms of spectral parameters, that is, frequency, amplitude, and phase.

First, an approximation of the bone-tissue signal spectrum for the vowel [i] has been reported (*3*) not to be consistent with the air-conduction signal spectrum.

Second, head resonance would be expected to affect the amplitude of bone-tissue sidetone components, a factor that is nonexistent via air conduction.

Third, a complex air-conduction signal cannot be assumed to be phase distorted in transmission, in view of the fact that the degree to which the velocity of sound is dependent on frequency in air is negligible. On the other hand Franke (*14*) found the propagation velocity of skull vibration to be a function of frequency. Additionally, whereas air-conduction sidetone is transmitted by means of compressional wavefronts, bone-tissue sidetone can be thought of as transmitted by both shear and compressional wavefronts or by a combination of the two if surfaces are involved in the pathways (*16*). In this respect it is important to consider the difference of propagation velocity inherent in the different types of propagation. Still another factor should be considered as conducive to phase distortion of the bone-tissue signal. Békésy (*6*) and Franke (*14*) established that vibra-

[1]As formulated by Husson (*17, 18*), the neurochronaxic theory of phonation purports that the vibration of the vocal cords is directly controlled by rhythmic neural activity that originates in the encephalon and is transmitted by the recurrent laryngeal nerves so that the muscle fibers of the larynx respond to the impulses 'coup pour coup.' As further implied, the pulmotracheal bellows as well as the approximated vibrating cords are required for sound production (*26*).

tion of the skull results in nodal lines of compression[2] of the skull as a function of frequency. The data of Franke showed a nodal line close to the anatomical position of the cochleae at 300 cps and nodal lines anterior to the cochleae at all higher frequencies. This is of interest when it is realized that a phase reversal of the signal can be expected across any such nodal line.

Fourth, the fact that a pure tone heard by bone-tissue conduction can be cancelled completely by phase and amplitude adjustments of simultaneously presented pure tones of the same frequency by air conduction has been utilized by many researchers. The additional fact that cancellation of complex tones cannot be attained with broadband phase shifters (*11*) would tend to show that the spectrum of bone-tissue sidetone is not comparable to that of air-conduction sidetone.

The foregoing context illustrates that phonation as well as speech per se is affected or controlled by audition concurrent to the activity. If the 'control signal' can be assumed to be a composite of bone-tissue and air-conduction sidetone transduced to the central nervous system then facts related to the pathways of sidetone would tend to indicate that the composite signal is possibly an aberration of the actual acoustic event. If this be the case, the extent of aberration inherent in the normal speaker in a conventional acoustic environment speculatively must fall within the limiting conditions of the auditory control system whereas certain experimentally induced aberrations would be seen as exceeding the limits of the control system as shown by quantifiable variations. In like manner, it is tempting to think that certain difficulties of speech are due to inherent aberration that exceeds the limits of the control system.

[2]The entire skull is displaced as a rigid body when the front portion is vibrated at low frequencies. At higher frequencies the front and back portions of the skull are displaced in opposite directions, that is, the displacement of the back portion, due to its inertia, does not parallel that of the front portion as the more elastic middle portion is compressed. The compressed portion does not exhibit displacement and is referred to as the nodal line of compression.

The Problem

Any interference with the production of speech or phonation, whether artificially produced or neuromuscularly inherent, is provocative to the student of speech science. This is shown, in part, by the abundance of confirming and expanding literature with regard to the effects of delayed sidetone. In contrast is the paucity of literature concerning experimentally induced blockage of phonation, except in a contraindicative vein (*2*). The present paper was a result of the author's attempt to duplicate the induced blockage of phonation.

Early in the attempt it became apparent that the effects of altering the phase of the vowel signal heard by air conduction were not easily discernible even with 'high amplification' unless the vowel signal was distorted, for example, by improper use of the line preamplifiers. With a nondistorted falsetto-vowel signal fed back at a 100-db sound-pressure-level output of the earphones, a vowel flutter was noticed and bodily sensations were reported inconsistently by the experimental subjects. This was in contrast to the highly consistent stoppage or blockage of phonation and perceived bodily sensations

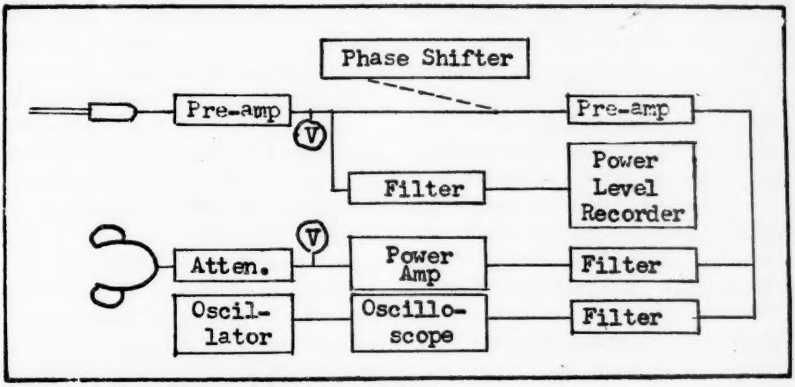

FIGURE 1. A block diagram of the instrumentation.

provoked by distorting a vowel sustained in falsetto and feeding it back to the ears at relatively low sound pressure levels in phase to that of the emitted vowel. The latter fact along with the suspected aberration of the normal sidetone experience served as motivation for the investigation.

The primary purpose of the investigation was to determine the influence of air-conduction sidetone spectrum distortion on the phonation of a sustained vowel. A secondary aim was to probe the relationship of phase and distortion of the signal to the experimental effect of phonatory blockage. In addition, the reactions of the subjects to the experimental effects were noted.

Instrumentation

A block diagram of the instrumentation is shown in Figure 1. The emitted vowel was picked up by a six-inch probe tube and transmitted via a condenser microphone system (Altec, Model M14) to a pair of series-connected preamplifiers (Presto, Model 40-B), the second of which was used to control the percentage of signal distortion. The output was filtered (Allison, Model 2AR), power amplified (Western Electric, Model 124), attenuated (Hewlett-Packard, Model 350B), and terminated on receivers (Permaflux, Model PDR3). The receivers were calibrated to be in phase with each other as well as with the input signal by means of an artificial ear (Bruel-Kjaer, Model 4109) and a phase angle meter (Acton, Model 320-AB). The receivers were mounted in extremely tight fitting muffs (David Clark, Model 372-8C-M) to prevent feedback from the receivers to the microphone.

An oscilloscope and a voltmeter enabled the subjects to monitor visually the fundamental frequency and the sound level of their sustained vowels. The amplified output of the microphone was filtered (Allison, Model 2AR) and formed a circular Lissajous figure on the oscilloscope when it matched the output frequency of an audio-signal generator (General Radio, Model 1304-B). The sound pressure level was indicated by a vacuum-tube voltmeter placed across the output of of the first preamplifier. A dual-beam

oscilloscope was used by the experimenter to monitor the waveforms at the outputs of the first preamplifier and the power amplifier. The output signal of the attenuator was monitored and controlled on the basis of a vacuum-tube voltmeter across the line at that point. A power level recording (Sound Apparatus, Model HPL) was made from the output of the first preamplifier through a band-pass filter (Krohn-Hite, Model 310-A) that excluded high-frequency breath noise.

A second part of the investigation necessitated an audio-signal delay unit and an electronic phase shifter. The design of the delay unit enabled the reproduce head to be minutely positioned in a continuous manner relative to the record head along the continuum that represented .02- to .25-sec delay. The electronic phase shifter was continuously variable through 360 degrees at either of two resonant points, 150 and 350 cps. At either frequency the amplitude variation was less than 1 db through 360 degrees and the effective phase shift was less for frequencies lower than and greater for frequencies higher than the resonant frequency.

Procedure

Eight normal-hearing male college students who served as experimental subjects in the first portion of the study were requested to practice sustaining the vowel [u] in falsetto at a fundamental frequency of 350 cps and at the lowest sound pressure level conducive to continuous phonation. Subsequent to the practice an identical procedure was followed as each subject experienced two conditions in each of four sessions, the rank order of conditions being reversed for each succeeding session. A minimum of two hours separated all sessions.

The diaphragm of the microphone was positioned 13 cm from the incisors with the probe tube extending into the oral cavity approximately 2 cm. Without the earphones on, the subject established and monitored the fundamental frequency of his vowel by a circular Lissajous figure on the oscilloscope. The lowest sound pressure level of the vowel that permitted continuous phonation was noted in terms of voltage and the needle position of the meter was used throughout the session as a monitor level by the subject. When the level and frequency of the vowel were established, the attenuator and the pen amplitude of the power level recorder were adjusted. The earphones were placed on the subject in a systematic manner and the session initiated without the subject knowing which condition was to be first. The subject signaled throughout the periods required for replenishing his breath supply and the power level recording was marked accordingly.

The two experimental conditions differed in terms of signal only, one having no appreciable nonlinear distortion, the other having 62.5% nonlinear distortion.[3] Both signals were

[3]A 350-cps sinusoidal signal was used as the input signal to be distorted by the system for calibration purposes. A frequency analysis of the distorted signal (Bruel-Kjaer Audio Spectrometer, Type 2109) disclosed that the second, third, and fourth harmonics were 8, 11, and 21 db below the level of the fundamental frequency, respectively. The percentage of distortion was calculated as the ratio of the energy in all the higher harmonics except the fundamental to the energy in all the harmonics including the fundamental frequency.

band-pass filtered to exclude frequencies below 75 cps and above 2040 cps. A sound pressure level of 82 db (re 0.0002 microbar) was maintained as the output of the earphones for the distorted signal whereas 98 db was maintained for the nondistorted signal. The difference in level was obtained subjectively by a loudness balance technique on 10 normal-hearing ears.

Excluding the phonation on the first breath and the time necessary for breath supply, the power level recording representative of 60 sec of phonation was scored for each condition using two criterion measures: (a) the number of phonatory blockages, and (b) the total time consumed by the blockages. By definition, a blockage of phonation occurred when the power level recording line indicative of the vowel amplitude fell to the baseline of the paper, that is, a fall of approximately 40 db sound pressure level. In the case of 'total-time-consumed-by-blocks' the distance between the fall-off of amplitude mark and the rise of amplitude mark was measured along a line 5 db below the maximum amplitude of the phonated vowel as indicated on the power level recording.

The results reported in Part A of the next section were obtained under the conditions described in the foregoing paragraphs. Part B of the next section describes results obtained by modifying the stated procedure in order to study the following: (a) the direct effect of the distorted sidetone condition using a population of subjects who had neither practiced sustaining the vowel in falsetto nor attempted phonation under the condition of nondistorted sidetone; (b) the effect of other distorted sidetone sound pressure levels; (c) the effect of distorted sidetone as subjects sustained the vowel [u] at their normal fundamental frequency rather than in falsetto; (d) the effect of distorting the sidetone of vowels other than [u] as well as certain voiced consonants sustained by subjects in falsetto; and (e) the effect of altering the phase of the fundamental frequency of distorted sidetone. In addition the results of sound spectrographic analysis of tape recordings made under the various conditions are included in Part B.

Results

Part A. Analysis of the pooled data by t tests for related means established that a significantly larger number of phonatory blockages ($t = 8.89$; $df = 7$; $t_{.01} = 3.499$) and greater total time consumed by blockages ($t = 11.79$; $df = 7$) occurred under the condition of distorted sidetone. The differences between the mean values for the paired observations in the four sessions for each subject were used as scores. Inspection of the original data disclosed that a mean of 5.84 blockages (ranging from 2.75 to 12.75 blockages) totaling 1.65 sec (0.30 to 3.65 sec) occurred under the condition of nondistorted sidetone whereas a mean of 40.15 blockages (26.25 to 61.25 blockages) totaling 17.67 sec (10.50 to 25.23 sec) occurred under the condition of distorted sidetone.

Part B. Direct Effect. The direct effect of the above described condition of distorted sidetone was subjectively evaluated for 32 subjects; 20 male college students, eight female college students, and four noncollege males, two of them approximately 50 years of age, one nine, and the fourth 11. In 28 cases

it was reported by the subjects and confirmed by the level recordings that phonatory blockages occurred whereas only inconsistent effects occurred for one male and two female college students and for one of the older subjects.

Evaluations by the subjects with regard to the experienced blockage revealed the following: (a) nearly all were aware of tension at the site of the larynx, suggestive of vocal-cord closure; (b) more than 75% of the subjects mentioned 'shortness of breath' or factors thought to be related to disturbance of breath control; (c) approximately 50% mentioned visceral tensions; (d) approximately 50% used extraneous body movements in attempts to regenerate the vowel after it had stopped; and (e) five subjects mentioned uncontrollable rapid lip tremors. False vocalization[4] and extremely high-pitched vocalization were elicited commonly by the subjects, yet for the most part they were unaware of the deviations.

Intensity of Air-Conducted Sidetone. The effect of varying the intensity of distorted sidetone was studied for five practiced subjects who were instructed to sustain the falsetto vowel in a passive yet continuous manner. At a 100-db sound pressure level as produced by the earphones the phonatory blockages were of a longer duration. In many cases the subject was unable to regenerate the vowel without stopping for breath supply. Initiation of a vowel was usually successful but phonation was of short duration before the cycle repeated itself. Lower levels comparable to that of hearing by bone-tissue sidetone (60 to 65 db, arrived at by loudness balance techniques) produced inconsistent effects, indicative perhaps of the threshold of the phenomenon based on the fact that a 5-db increment would provoke the effect anew but to a lesser degree than at higher levels.

Fundamental Frequency of the Vowel. Five practiced subjects were instructed to phonate [u] at their natural fundamental frequency and at the lowest sound pressure level conducive to continuous phonation. With a distorted signal of 85 db as produced by the earphones the phonation of three subjects was blocked in a consistent manner. The fourth subject experienced occasional phonatory blockages whereas the remaining subject showed inconsistent effects only. Five unpracticed subjects were instructed as stated above. In every case the vowel was evaluated as hoarse in quality, and false vocalization was in evidence throughout the sessions. With increased attention toward maintaining a somewhat stable fundamental frequency, four of the five subjects experienced phonatory blockages of short duration. It was the consensus among those subjects who experienced the phenomenon in falsetto as well as in this condition that the effect was provoked to a greater degree when the vowel was sustained in falsetto than in this condition.

Effect on Other Signals. The effect of sidetone distortion on signals other than [u] was observed in a preliminary manner on five practiced subjects, that is, the sound pressure level, per cent of

[4]Continuous vocalized 'sputtering' with a fundamental frequency about one octave lower than that of preceding and succeeding vocalization is called here *false vocalization.* This nonsemantic acoustic output is also called ventricular phonation and has been attributed to activity of the false vocal cords (*32*). (See Figures 3, 4, and 5.)

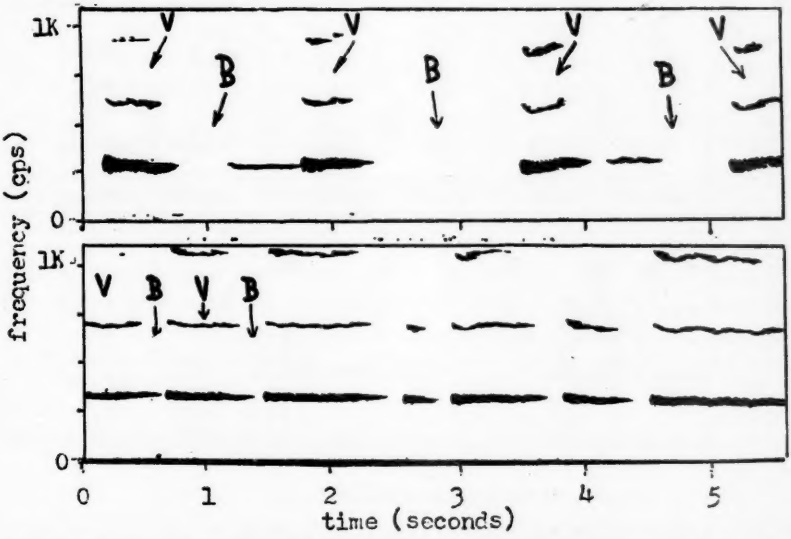

FIGURE 2. Two sound spectrograms with frequency as the ordinate, time as the abscissa, and intensity indicated by the density of the marking: (upper) sustained falsetto vowel [u] blocked by voluntarily closing the vocal cords; (lower) sustained falsetto vowel [u] blocked by distorted sidetone. V = vowel, B = blockage of phonation.

distortion, and the phase of the fundamental frequency of the signal were varied and confounded. Under these conditions sidetone distortion was observed to provoke phonatory blockage on the following signals sustained in falsetto: [o], [ɔ], [æ], [ɛ], [ɪ], [i], [m], [l], [b], and [g]. The lips were pursed around the probe tube for [m] and [b].

Alteration of Phase. Without distortion of the signal only inconsistent effects were shown as the phase of the fundamental frequency of the vowel was shifted through 360 degrees. Approximately the same effect obtained when the band-pass filter was set to include only the 350-cps fundamental frequency of the vowel as the fundamental frequency was shifted in phase. When the signal was distorted and the band width was 75 to 2040 cps, the effect was obtained through the range of 360 degrees but to a lesser degree than that obtained without the phase shifter in the line. The latter fact could have been due to the filtering action inherent in the resonant-type phase shifter. One other subjective observation would seem worth mentioning. As the phase control dial was rotated slowly, the phonatory blockage would become a much more dramatic event on a spasmodic basis. If the control was stopped at such a point the blockage would persist in its aggravated form, but if the dial was turned swiftly from the point, phonation would return only to be blocked again when the phase control was repositioned. The phase reading was not stable from one breath supply to the next and a period

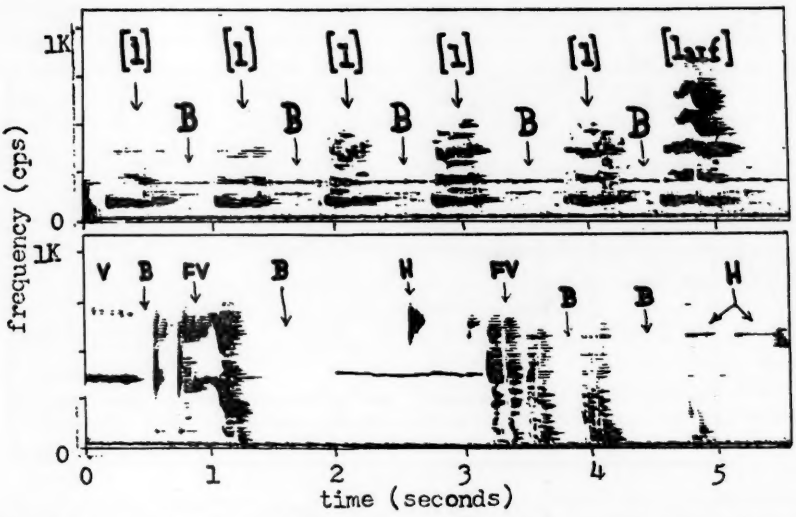

FIGURE 3. (upper) A clonic-type stuttering block on the word *life*. (lower) Sustained falsetto vowel [u] blocked by distorted sidetone, the fundamental frequency of which was shifted in phase showing extreme pitch variations as a function of time. V = vowel, B = blockage of phonation, H = high-frequency vocalization, FV = false vocalization.

of tabulating such points for given subjects was not enlightening.

A second attempt to establish the phase relationship utilized the delay unit mentioned previously.[5] This method was used because it provides, within practical limits, a faithful reproduction of the signal spectrum, that is, the amplitude and the phase of the components are preserved as a function of their frequencies. In this case however the vowel was not heard within its first wave length due to the minimum delay and the tape speed (60 in. per sec) of the instrument. With this fact in mind it was observed that neither a nondistorted nor a distorted signal presented at equal loudness levels over a range of sound pressure levels caused consistent effects as the delay was varied between 0.02 and 0.25 sec. Little or no alteration of effect was noticed as a result of reversing the polarity of both earphones simultaneously or one earphone as compared to the other in any of the conditions of Parts A and B.

Sound Spectrographic Analysis. Sound spectrograms (Kay Electric, Model 651-A) were made from tape

[5]Conceivably, the delay in time may be viewed as a phase shift of the air-conduction sidetone experience to that of the bone-tissue sidetone experience or to the original signal. Using a complex nonperiodic signal such as connected speech, a cycle representing 360 degrees of phase shift would be the period of a phoneme, syllable, word, or phrase, depending on one's point of view. As the delay period exceeded the signal period the phase shift would be within or exceed 360 degrees. In the case of a (pseudo) periodic signal such as a sustained vowel, a cycle representing 360 degrees of phase shift would be dependent on its periodicity and a phase shift or discrepancy of more than 360 degrees would not be theoretically possible, as in the case of sinusoidal signals.

recordings procured under conditions similar to those of Part A. The same eight subjects blocked phonation of the vowel [u] by approximating the vocal cords while hearing the nondistorted signal whereas in the second condition phonation was blocked by the distorted sidetone. The consensus of five people who judged the spectrograms was that the frequency bars of the voluntarily blocked vowels appeared to swing upward (increase of frequency) just prior to the onset of blockage as compared to the nearly straight extension of the bars of the involuntarily blocked vowels at the onset of blockage. Spectrograms displaying these phenomena for one subject are shown in Figure 2. The straight extension of the bars has been observed also in several recorded samples of clonic-type stuttering blocks, one of which is shown in Figure 3. It is of interest to note that Rawnsley and Harris (27) made a similar observation on syllables or sounds repeated by nonstutterers under the condition of delayed sidetone. Figure 3 displays also a sustained falsetto vowel [u] blocked by distorted sidetone, the fundamental frequency of which was shifted in phase, showing extreme pitch variations as a function of time.

Vannier and others (*31*) mentioned that induced blockage was not observed when the vowel was sustained in chest register. Representative spectrograms displayed in Figure 4 show examples of the vowel [u] sustained at a normal fundamental frequency of voice under the condition of nondistorted sidetone with no occurrence of blockage and the vowel [u] at normal fundamental blocked by distorted sidetone showing concomitant false vocali-

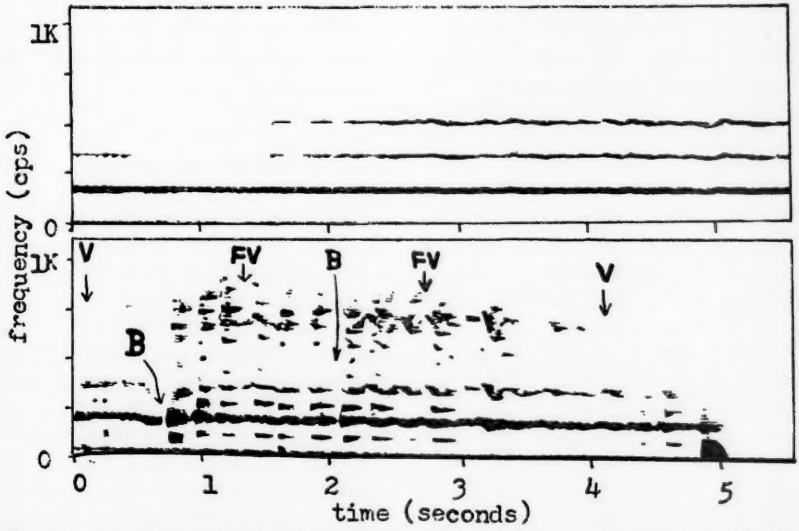

FIGURE 4. (upper) A representative spectrogram of the vowel [u] sustained at a normal fundamental frequency of voice under the condition of nondistorted sidetone. (lower) Sustained vowel [u] at normal fundamental frequency blocked by distorted sidetone showing concomitant false vocalization. V = vowel, B = blockage of phonation, FV = false vocalization.

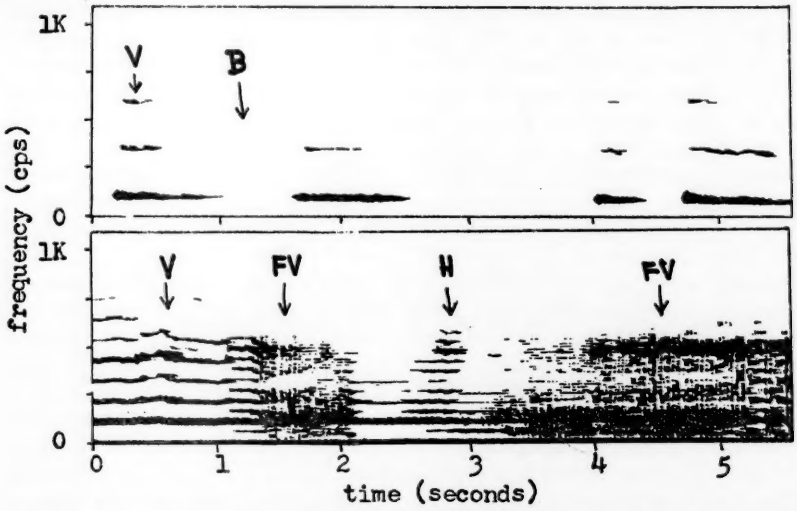

FIGURE 5. (upper) Sustained vowel [u] at normal fundamental frequency blocked by distorted sidetone. (lower) Sustained vowel [u] at normal fundamental frequency under the condition of distorted sidetone showing extreme pitch variations as a function of time. V = vowel, B = blockage of phonation, H = high-frequency vocalization, FV = false vocalization.

zation. The spectrograms displayed in Figure 5 show examples of induced blockage and also extreme pitch variations as the vowel [u] was sustained at a normal fundamental under distorted sidetone.

The results of using nonfiltered recordings and increasing the frequency range of the spectrograms to 8000 cps showed that the intervals of blockage were devoid of high-frequency markings, which, if evidenced, would have indicated that the breath stream was traversing the oral cavity. This is substantiated somewhat by the fact that the pickup probe extended 2 cm into the oral cavity during recording.

Discussion

The results of this study are difficult to compare to those reported by Vannier and others (*31*) because of their omission of details. Whereas they stated merely that the emitted vowel was picked up by a microphone, microphone placement in the preliminary stage of this study was varied from 5 cm within the oral cavity to 13 cm directly in front of the incisors, with and without the probe tube, with no apparent phonatory blockage through a 360-degree phase shift irrespective of signal band width or level. The placement used throughout the present study seemingly allowed a normal delay of the air-conduction sidetone signal when it is considered that the signals emitted by the earphones were in phase with the signal following the microphone. Under these conditions phonatory blockage, although it occurred only with signal distortion, seemed to be affected by the phase of the fundamental frequency.

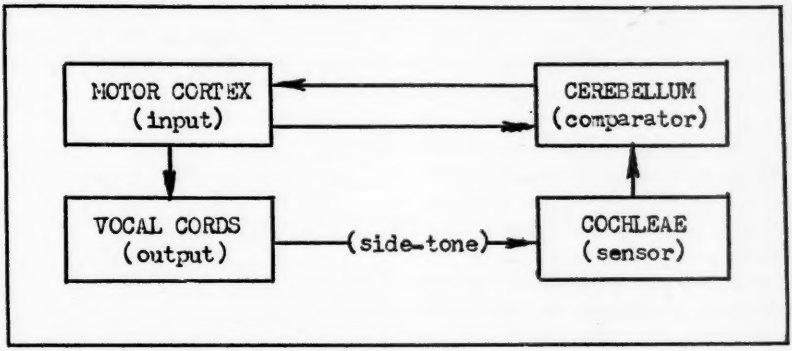

FIGURE 6. A block diagram of a theoretical sidetone control circuit.

Without implication, mention should be made of two possible sources of distortion in the work of Vannier and others (*31*). From the description of their phase shifter it could be assumed that the unit was resonant at 300 cps and as such it effected a differential phase shift for the components of the vowel spectrum, especially since filters were not mentioned as line equipment. A second possibility of distortion would be the effect of 'high amplification,' this based on the inability of electronic amplifiers to handle more than a certain signal amplitude without introducing some distortion of its form or spectrum.

On the basis of the foregoing reservations the experimental blockage of phonation obtained under the conditions of this study may be viewed as evidence confirming the results reported by Vannier and others (*31*). The results of spectrographic analysis of voluntarily and involuntarily provoked phonatory blockages would tend to lend credence also to their stroboscopic observation of a medially-occluded position of the vocal cords at the time of blockage.

Tentative Explanation of the Blockage. Utilizing established neural pathways, a theoretical point of view may be taken based for the most part on the following by Ruch (*28*):

> '. . . the cerebellum could be likened to the comparator of a servomechanism. It may receive from the cerebral cortex some representation of the command and from the muscles a representation of the resulting movement. These, compared, may result in a signal that is transmitted to the motor cortex, altering its signal to the muscles so as to diminish the error. . . . However, the cerebellum receives impulses from the distance receptors, and from the cortical areas for vision and audition. These pass to the neocerebellum which is especially concerned with voluntary movement. Such an arrangement may mean that the cerebellum can compare the true input, the auditory or visual representation of the goal, and the proprioceptive indication of the limb's position. . . .'

A modification of the foregoing possibility would be an input-informed feedback circuit in which the auditory representation was transmitted to the cerebellum from the cochlear nuclei (*15*) rather than along the more circuitous route to the auditory cortex, cerebellum, and then back to the motor cortex. With reference to Figure 6,

the corticifugal impulses necessary for voluntary phonation, in addition to activating the laryngeal musculature and other related systems via corticobulbar and corticospinal pathways, would be transmitted to the cerebellum by way of frontopontocerebellar pathways. Considering the auditory representation of the goal (phonation), the composite bilateral signals of air-conduction and bone-tissue sidetone would be transduced on the basilar membranes and transmitted via the cochlear nuclei to the cerebellar auditory centers which in turn are connected to the (pre) motor cortex by way of the cerebellar cortex, dentate nuclei, and the thalamus. In this manner the 'report back' could be conceived of as occurring with sufficient rapidity to permit control within the course of phonation of a vowel in context.[6]

Speculatively, the manner in which the output signal would be scanned by such a system would be that the amplitude and periodicity of the signal could provide the bits of information necessary for comparison to the input signal with subsequent control of the input signal based on the discrepancy. To digress, it was mentioned that blockage of phonation occurred significantly more when a distorted vowel was the signal. When the signal was not distorted the spectrum was dominated by the fundamental frequency whereas the amplitude of the second harmonic added by the nonlinear distortion was of the same order of magnitude as the fundamental.[7]

If, as stressed by many theories of hearing, analysis of a complex signal occurs at the level of the basilar membrane, the distorted signal would produce centripetal impulses representative of the harmonics inherent in the signal. If no one harmonic served as the predominant bit of information provided the cerebellum (comparator) it is conceivable that the system would be thrown into a hunting type of oscillation, which in turn could be reflected in wide swings of the fundamental frequency of phonation or a breakdown of the system.

The phonatory blockage and the concomitant reactions could be viewed as satisfying at least in part the three cardinal signs of cerebellar disorder as listed by Ruch (*28*): (a) disorders in the direction, rate, range, and force of movement; (b) tremor; and (c) hypotonia. With regard to the latter, the observations by Vannier and others (*31*) and the interpretations of spectrograms in this study tentatively show that the vocal cords were arrested in a position of adduction at the time of blockage. Moulonguet, Laget, and Husson (*23*) previously reported that the action potentials of the recurrent nerves

[6]An average value for the duration of a vowel in context, as calculated from Black (7) and Fletcher (*13*), is 130 msec. This is to be compared with the shortest latency time of 10 msec between the onset of a tone and the resulting cortical response mentioned by Galambos (*15*). Although it cannot be assumed that this is the case via the cochleocerebellocortical pathway, the difference in time seemingly lends credence to the possibility of control during performance.

[7]Frequency analyses of eight nondistorted and eight distorted falsetto [u] signals disclosed that the mean second, third, and fourth harmonics were 10.7, 18.8, and 24.2 db below the mean level of the fundamental frequency for the nondistorted signals and 4.4, 10.5, and 13.3 db for the distorted signals. The eight pairs of signals were recorded from the eight subjects who served in the first portion of the study.

were synchronous to the abduction of the vocal cords. On this basis it would seem logical to assume that the abductor muscles were not fired off at the time of blockage or, said another way, they could be viewed as having been in a state of hypotonia.

Implications of the Blockage. The lack of basic information concerning the organism at the time of experimental blockage of phonation makes it presently impossible to discuss its implications for dealing with clinical-type speech disorders, especially when the lack of similar information concerning the latter is appreciated. Yet, three effects of distorted sidetone were observed consistently enough to indicate their possible importance in terms of further research: (a) blockage of phonation; (b) extreme shifts of the fundamental frequency of phonation; and (c) voice quality of an undesirable nature.

Although mentioned by some writers, the relationship of phonation to such early symptoms of stuttering as repetitions and prolongations has not been studied systematically. In many cases it can be observed clinically that the repetitions are characterized by noninitiated (that is, omitted) vowels or voiced sounds that should be included in the syllable attempted or by vowels or voiced sounds that are initiated but not completed in the sense that transition to the following sound is accomplished. In other cases the prolongations are characterized by shifts of the fundamental frequency of vowels or voiced sounds, for example, the elevation of pitch as the sound is prolonged in transition to the following sound. On this basis it could be hypothesized that blockage or deterrents of phonation in general serve as the locus of difficulty in at least a portion of the stuttering population.

In other studies, it has been observed (a) that when stutterers were unable to hear their voices (sidetone) due to auditory masking they were less likely to stutter (*24,29,10,21*), (b) that whispering showed a marked tendency to be associated with a decrease in stuttering (*1*), and (c) that the use of an electrolarynx (instead of vocal-cord activity) resulted in approximately a 70% reduction of stuttering over a three-week period (*22*). Inasmuch as phase of sidetone was observed by Vannier and others (*31*) to block phonation and was observed in this study to control the blockage, it is of interest to note that stutterers as a group exhibited an interear disparity with regard to the phase of bone-tissue sidetone (*30*).

The experimental effects of extreme shifts of the fundamental frequency of phonation and voice quality of an undesirable nature could be hypothesized as related to the findings of Bowler and Anderson (*9*) in which a fundamental frequency analysis of harsh voice quality disclosed abrupt and extreme changes of fundamental frequency occurring at an average of 1.6 times per second within the lower half of the frequency range of individual subjects.

Summary

The primary purpose of the investigation was to determine the influence of air-conduction sidetone signal distortion on the phonation of a vowel sustained in falsetto. Subjects experienced two experimental conditions that

differed in terms of signal only, one having no appreciable nonlinear distortion, the other having 62.5% nonlinear distortion. The results and conclusions were as follows: (a) Under the conditions of the study phonatory blockage occurred significantly more often when the air-conduction sidetone signal was distorted. (b) Under the conditions of the study a significantly greater total time was consumed by the phonatory blockages when the air-conduction sidetone signal was distorted. (c) Sound spectrographic analysis revealed that essentially the same pattern existed when phonation was blocked by voluntarily approximating the vocal cords as when phonatory blockage was provoked by the distorted air-conduction sidetone signal. (d) Three effects of air-conduction sidetone signal distortion were observed to a degree of consistency that would stress their possible importance in terms of further research: blockage of phonation; extreme pitch variations of phonation; and voice quality of an undesirable nature. (e) With certain reservations the experimental blockage of phonation obtained under the conditions of this study may be viewed as confirmative evidence to the results reported by Vannier and others (*31*).

Acknowledgment

The author wishes to acknowledge the technical assistance provided by William L. Dawson, Technician, Department of Speech, Ohio State University.

References

1. Adams, Lois H., A comparison of certain sound wave characteristics of stutterers and nonstutterers. In W. Johnson (Ed.), *Stuttering in Children and Adults*. Univ. Minnesota Press, 1955.
2. Arslan, M., Experimental research on the cochleo-recurrential reflex. An *Abstracts: International Voice Conference*. Evanston: Northwestern Univ., 1955.
3. Atkinson, C. J., An estimate of the spectrum of body-conducted side-tone for the vowel [i]. Univ. Iowa, Signal Corps; Cont. DA-36-039 sc-63144, *Supplementary Report 3*, 1955.
4. Baccaro, P. M., A comparative study of cortical potentials of stutterers and nonstutterers under normal reading and delayed speech feedback conditions. Unpublished master's thesis, Univ. Mississippi, 1955.
5. Békésy, G. v., The structure of the middle ear and the hearing of one's own voice by bone conduction. *J. acoust. Soc. Amer.*, 21, 1949, 217-232.
6. Békésy, G. v., The mechanical properties of the ear. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley, 1951.
7. Black, J. W., The quality of a spoken vowel. *Archives of Speech*, 2, 1937, 7-27.
8. Black, J. W., The effect of delayed side-tone upon vocal rate and intensity. *J. Speech Hearing Dis.*, 16, 1950, 56-60.
9. Bowler, N. W., and Anderson, V. A., A fundamental frequency analysis of harsh vocal quality. In *Abstracts: International Voice Conference*. Evanston: Northwestern Univ., 1955.
10. Cherry, C., Sayers, B., and Marland, Pauline, Experiments upon the total inhibition of stammering by external control, and some clinical results. *J. Psychosom. Res.*, 1, 1956, 233-246.
11. Dolch, J. P., and Schubert, E. D., Study of body-conducted side-tone. Univ. Iowa, Signal Corps; Cont. DA-36-039 sc-42562, *Supplementary Report 6*, 1954.
12. Fairbanks, G., and Guttman, N., Effects of delayed auditory feedback upon articulation. *J. Speech Hearing Res.*, 1, 1958, 12-22.
13. Fletcher, H., *Speech and Hearing in Communication* (2nd ed.). New York: Van Nostrand, 1953.
14. Franke, E. K., Response of the human skull to mechanical vibrations. *J. acoust. Soc. Amer.*, 28, 1956, 1277-1284.
15. Galambos, R., Neural mechanisms of audition. *Physiol. Rev.*, 34, 1954, 497-528.

16. Gierke, H. V., Oestretcher, H. L., Franke, E. K., Parrack, H. O., and Wittern, W. V., Physics of vibrations in living tissue. *J. appl. Physiol.*, 4, 1952, 886-900.
17. Husson, R., Étude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée. Thèse Fac. Sci. Paris, 17 Juin 1950, *Rev. sci., Paris*, 3.306, 3.307, 3.308, 1950 (couronnée par l'Institut).
18. Husson, R., Physiol. phonat. du larynx, *Encycl. Méd.-Chirurg.*, Vol. O.-R.-L., 20,632 B 10, Paris, Nov. 1955.
19. Lee, B. S., Effects of delayed speech feedback. *J. acoust. Soc. Amer.*, 22, 1950, 824-826.
20. Lee, B. S., Artificial stutter. *J. Speech Hearing Dis.*, 16, 1951, 53-55.
21. Maraist, Jean Ann, and Hutton, C., Effects of auditory masking upon the speech of stutterers. *J. Speech Hearing Dis.*, 22, 1957, 385-389.
22. McKenzie, F. A., A stutterer's experiences in using an electrolarynx. In W. Johnson (Ed.), *Stuttering in Children and Adults*. Univ. Minnesota Press, 1955.
23. Moulonguet, A., Laget, P., and Husson, R., Démonstration, chez l'homme, de l'existence dans le nerf récurrent de potentiels d'action moteurs synchrones avec les vibrations des cordes vocales. *Bull. Acad. nat. Méd.*, 137, 1953, 475-482.
24. Paruzynski, T., A study of the speech behavior of stutterers using the influence of circumstances created by inhibiting auditory and visual sensations. Unpublished master's thesis, Marquette Univ., 1951.
25. Peters, R. W., The effect of changes in side-tone delay and level upon rate of oral reading of normal speakers. *J. Speech Hearing Dis.*, 19, 1954, 483-490.
26. Portmann, G., Physiology of phonation. In *Abstracts: International Voice Conference*. Evanston: Northwestern Univ., 1955.
27. Rawnsley, A. L., and Harris, J. D., Comparative analysis of normal speech and speech with delayed sidetone by means of sound spectrograms. BuMed., Surg. Res. Proj. NM 003 041.56.03, Med. Res. Lab., New London, *Report No. 248*. 1954.
28. Ruch, T. C., Motor Systems. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley, 1951.
29. Shane, Mary Lou S., Effect on stuttering of alteration in auditory feedback. In W. Johnson (Ed.), *Stuttering in Children and Adults*. Univ. Minnesota Press, 1955.
30. Stromsta, C. P., A methodology related to the determination of the phase angle of bone-conducted speech sound energy of stutterers and non-stutterers. Unpublished doctor's thesis, Ohio State Univ., 1956.
31. Vannier, J., Saumont, R., Labarraque, L., and Husson, R., Production experimentale de blocages synaptiques recurrentiale par des stimulations auditives homorythmiques avec dephasages reglables. Bordeaux: *Revue de Laryngologie*, 1954.
32. Van Riper, C., and Irwin, J. V., *Voice and Articulation*. Englewood Cliffs, N. J.: Prentice-Hall, 1958.
33. Wiener, N., *Cybernetics*. Cambridge: Technology Press, 1948. (Also New York: John Wiley, 1948.)